Journal of Development Economics 144 (2020) 102418

Contents lists available at ScienceDirect

Journal of Development Economics

journal homepage: www.elsevier.com/locate/devec

Cash-plus: Poverty impacts of alternative transfer-based approaches

Richard Sedlmayr[a,*], Anuj Shah[b], Munshi Sulaiman[c]

[a] University of Oxford, United Kingdom
[b] University of Chicago, United States
[c] BRAC Uganda, Uganda

ARTICLE INFO

JEL classification:
O12
O22
I38

Keywords:
Graduation
Microenterprise
Cash transfers
Behavioral design

ABSTRACT

Can training and mentorship expand the economic impact of cash transfer programs, or would such extensions waste resources that recipients could allocate more impactfully by themselves? Over the course of two years, a Ugandan nonprofit organization implemented alternative poverty alleviation approaches in a randomized manner. These included an integrated graduation-style program involving cash transfers as well as extensive training and mentorship; a slightly simplified variant excluding training on savings group formation; and a radically simplified approach that monetized all intangibles and delivered cash only. Light-touch behavioral extensions involving goal-setting and plan-making were also implemented with some cash transfer recipients. We find that simplifying the integrated program tended to erode its impact.

1. Motivation

So-called graduation programs provide an integrated package of interventions at no cost to exceptionally poor households in low-income countries. The package is comprised of transfers (in the form of productive and/or monetary assets) as well as training and coaching activities (that aim to cultivate some intangible asset). Randomized evaluations have found graduation programs to be cost-effective on average, though with considerable heterogeneity across contexts (Banerjee et al., 2015). Long-term (7–8 year) effects appear highly encouraging (Bandiera et al., 2017; Banerjee et al., 2017a).

The intangible (training and mentorship) component usually accounts for a substantial portion of program costs. Skeptics could argue that development practice has a long history of paternalistically misallocating resources (Bauer, 1972; Easterly, 2007; Moyo, 2009; Shapiro, 2017). Why invest in training and mentorship on participants' behalf? Why not give beneficiaries expanded agency over program resources by expanding the monetary transfer portion of the program and allowing them to invest for themselves?

One justification could be that human capital markets are inefficient or inaccessible (Stiglitz, 1989). For example, they may be disjoined from financial markets because human capital cannot be easily collateralized (Ljungqvist, 1993). This might explain why the graduation approach invests in the creation of savings devices. But the approach goes much further, providing extensive training and coaching on microenterprise development and beyond. Its design implies that if participants were left to their own devices – so provided only with unconditional cash transfers – they would under-invest in human capital against their own best interests.

The potential for such "internalities" is well established. Recipients' investment decisions can be sensitive to seemingly trivial design features such as the labeling of the transfer (Benhassine et al., 2015). Many other behavioral factors – ranging from time inconsistency (O'Donoghue and Rabin, 1999) to social preferences (DellaVigna, 2009) – might cause recipients to forgo profitable investment opportunities. Further, people's beliefs may be flawed: for boundedly rational agents who operate with imperfect information and finite attention, it can be a tall challenge to accurately perceive the returns to available investment options. Program implementers with access to objective data may be in a position to appraise returns to human capital more accurately than the individuals in question can (Jensen, 2010).

If integrated microenterprise development is in fact a more cost-effective poverty alleviation approach than unconditional cash transfers, this could have significant policy repercussions. In Sub-Saharan Africa alone, over US$ 25M are disbursed in the form of unconditional cash transfers on a daily basis (World Bank, 2019), and the composition

* Corresponding author.
E-mail address: richard.sedlmayr@gmail.com (R. Sedlmayr).

https://doi.org/10.1016/j.jdeveco.2019.102418
Received 24 April 2018; Received in revised form 13 November 2019; Accepted 13 November 2019
Available online 27 November 2019

of transfer programs could be altered by adding intangible (training and mentorship) extensions to the existing targeting and disbursement infrastructure.[1] Evidence on the relative impacts of integrated microenterprise programs and pure cash would therefore appear highly actionable. A broad literature on the impacts of unconditional cash transfers exists (Bastagli et al., 2016), but longer-term economic impacts are less well established (Wydick, 2018). While highly encouraging impacts have been found in contexts involving middle-income entrepreneurs (De Mel et al. 2012; McKenzie, 2017), the long-term impacts on the very poor appear more muted (Blattman et al., 2018; Haushofer and Shapiro, 2018). Some recent studies suggest that human capital extensions can help the poor derive more sustained value from cash or asset transfers (Banerjee et al., 2018; Berge et al., 2015; Chowdhury et al., 2017) though not all research agrees (Fiala, 2018; Karlan et al., 2015). The high degree of heterogeneity in available evidence suggests that impacts may be moderated by important unknown factors (Cartwright and Hardie, 2012; Deaton, 2010). If some of these include tacit ones that can neither be well defined nor monitored (say, some aspect of "implementation quality"), and these factors are negatively associated with scale, then pilot studies have yielded exceedingly optimistic policy predictions (Bold et al., 2013; Pritchett and Sandefur, 2014).

2. Research questions

We are interested in estimating the impacts of the integrated microenterprise development program, both in absolute terms and relative to plain cash transfers. Based on the available body of evidence, we expected that the integrated program would direct households towards microenterprise administration and lead to sustained improvements in economic as well as subjective well-being. Meanwhile, based on work by Fafchamps et al. (2014), we expected that unconditional cash transfers would result in relatively lower initial investment in productive assets, leading to high short-term but subsequently lower consumption impacts compared to the integrated microenterprise development approach.

Second, we seek clues for how to potentially simplify the integrated graduation approach; all else equal, an alternative with fewer and more templatized interactions between program implementers and beneficiaries would be more scalable. We can simplify the integrated approach by subtracting individual components. The theoretical justification for including savings devices in the integrated approach is somewhat ambiguous, and available evidence on savings groups as a stand-alone intervention is mixed (Gash and Odell, 2013; Karlan et al., 2017). The savings group component is therefore a plausible candidate for elimination.

The approach might be simplified further by identifying "key active ingredients" and delivering them in distilled form. The graduation approach has been shown to improve markers of subjective well-being (Banerjee et al., 2015), and it has been suggested that optimistic belief sets could have mediated – so been causally responsible for – the observed poverty reduction (Duflo, 2012). This raises the question if a light-touch intervention that directly targets psychological mechanisms could yield similar benefits to the integrated program at a fraction of the cost, or at least expand the benefits of pure cash transfers.

We also explore methodological innovation in the domain of research transparency. The richness of the data set creates room for a wide range of analytical approaches and thereby the problem of data mining: when many different analyses are pursued, but reporting focuses on those results that turn out to be significant, p values become biased and there is an elevated chance that apparent discoveries are in fact spurious (Ioannidis, 2005). So-called pre-analysis plans are one potential remedy: by committing to key aspects of the analysis before data are available, these can demonstrate that results were not cherry-picked (Casey et al., 2012). But pre-analysis introduces an epistemic trade-off, and the loss can exceed the gain (Olken, 2015). For example, pre-selected specifications may end up being statistically weak choices. We explore the use of alternative research transparency tools that allow making analyses contingent on the data. For this purpose, we first generate and visualize an entire universe of plausible results, then arrive at final inferences by restricting this universe in a traceable manner.

3. Programmatic context

Village Enterprise is a nonprofit organization that implements microenterprise programs in Uganda and Western Kenya. Like most other graduation programs, it uses a participatory targeting process as well as a proxy means test to identify the poorest households and then provides one of the household members with a combination of transfers, mentorship, and training. The program is relatively short for a graduation program: training sessions take four months, mentorship engagement takes nine months, and the overall program concludes within a year. A substantial part of the training is focused on microenterprise administration (e.g., business selection, business planning, record-keeping, and livestock management). The program encourages participants to establish business activities as partnerships with other households (target size: three households). The program also establishes village-level savings groups with about thirty members that provide basic savings and loan functions and train participants on the formation, functioning, and governance of these groups. There is little training beyond microenterprise and savings group formation; the program does not include modules included in diverse other integrated development programs, such as nutrition, hygiene, family planning, child rearing, or literacy. Unusually, the program does include a training session on environmental conservation. Coaching is run by designated business mentors and focuses specifically on matters of micro-enterprise management. The transfer component of the program is delivered not in the form of physical assets, but cash. Transfers are made to the business partnership, not to individuals or households, on the presumption that this will encourage productive investment. Indeed, the second of the two transfer instalments is made conditional on having invested the first instalment in the group business. Unlike in some comparable programs, no consumption stipend is provided. Being less comprehensive and shorter in duration, the Village Enterprise program comes at roughly a third of the cost (in USD PPP terms) of the least costly graduation program included in the meta-study of Banerjee et al. (2015, 2017b).[2]

4. Study design

4.1. Sampling and eligibility

Two regions were selected for the study – one in Western Uganda (Hoima district) and another in Eastern Uganda (Amuria, Katakwi, and Ngora districts). In each region, 69 villages were identified that qualified as large enough for the study, meaning that an initial mapping exercise indicated that at least 70 participant households would qualify for the Village Enterprise program.

In these villages, Village Enterprise implemented a participatory wealth ranking similar to the participatory rural appraisals described by

[1] For the programs to fully resemble the so-called graduation approach, many would further require a change in the payout pattern of transfers – specifically, the inclusion of larger and less frequent lump sum payments.

[2] If one was to interpret the programmatic imperative underlying graduation to involve "big push" investments, the low relative cost of the Village Enterprise program limits the generalizability of results to other graduation programs. Meanwhile, if one interprets the programmatic imperative to involve "multi pronged" investments, comprised of both tangible and intangible components; and if one does not assume economies of scale; then it is warranted to generalize rates of cost-effectiveness from the Village Enterprise program to costlier graduation programs.

Table 1
Economic Status of Eligible Households at Baseline (UGX per capita).

	Western Region		Eastern Region	
	Mean	Std Dev	Mean	Std Dev
Total Consumption (Annual)	**747,387**	**399,386**	**505,844**	**311,176**
Food & Beverage Consumption	586,012	316,141	378,748	252,055
Recurring Consumption	87,630	83,409	59,573	59,316
Infrequent Consumption	61,815	71,185	60,435	66,710
Total Net Assets	**102,197**	**114,654**	**95,196**	**111,343**
Livestock Assets	32,220	58,211	60,752	82,027
Durable Assets	62,752	64,674	30,869	39,711
Net Financial Position	1,830	8,187	833	6,557
Total Productive Cash Inflows (Annual)	**268,442**	**326,199**	**104,266**	**164,515**
Net Cash Inflows from Farming	11,128	67,611	−9,024	49,288
Income from Other Self-Employment	94,324	169,463	39,482	91,791
Income from Paid Employment	124,588	176,312	66,532	89,469

Notes: As data are derived from baseline survey, they are contingent on study recruitment and survey consent. All items are calculated in accordance with the preferred operationalization presented below, meaning that the top and bottom 2.5% of observations are coded to the respective cutoff levels (95% winsorization). As each measure is winsorized separately, the sum of the sub-aggregates may not equal the aggregates.

Banerjee et al. (2009), involving the active participation of village residents in the identification of ultra-poor households. Households ranked in the poorer two of four possible wealth categories were slated for further screening.

A basic household survey was then administered to these households with 10 survey questions (see Schreiner, 2011) that were used to calculate a Poverty Probability Index[3] (PPI). Households with poverty scores at or below 39 out of 100 were considered eligible, except if any of the following exclusion criteria applied: a) there is a teacher, salaried worker or pension recipient in the household; b) the household owns more than two cows; or c) the dwelling has a cement floor, brick wall, and metal roof. Households with poverty scores above 39 were considered ineligible, except if two or more of the following inclusion criteria applied: a) the household head is unemployed and not pensionable; b) the household has 8 or more children aged under 18; c) the household head is disabled, widowed, or an orphan under the age of 18; d) the household has suffered directly from a natural catastrophe; or e) the household head is living with a chronic illness.

Table 1 provides a sense of the economic status of eligible households at baseline. Per capita consumption averages UGX 617k annually, amounting to 1.59 USD PPP (2016) per day (for applicable rates, see Appendix A: Exchange Rates. It appears that the implementer successfully targets people whose consumption lies below the international poverty line of USD PPP 1.90 per capita per day. Our measures indicate that most consumption is not derived from income earned in the form of cash inflows from productive activities, which suggests that households derive a large share of consumption from subsistence or assistance. That said, income measures are notoriously difficult to measure in low-income contexts and especially prone to under-reporting (see Deaton, 1997; Meyer and Sullivan, 2003).

During the PPI survey process, Village Enterprise also identified a representative for each household who might be invited to receive a development intervention. The resulting list was shared with the research team for randomization.

4.2. Randomized assignment

The chosen randomization protocol was an effort to optimize across parallel objectives to implement impactfully, improve future operations, and expand the body of knowledge. The allocation of the study population to different arms and sub-arms was therefore determined not only by the scientific considerations of the research team but also by accommodating the needs and constraints of the implementer. The implementer was largely unconstrained in the availability of resources for the implementation of its standard program. The further removed a program variant, the harder it was for the implementer to justify the use of charitable resources. As a consequence, the program variant without the savings group component was assigned a smaller sample than the standard program. Pure cash transfers (which required dedicated evaluation resources) were implemented on an even smaller scale. Meanwhile, we included some variants that were closely associated with standard operations and could be funded by the implementer and assessed for operational research purposes but were implemented at too small a scale to warrant scientific attention. The analyses presented in this paper are those from which some positive scientific contribution was expected – but as the sample sizes allocated to different tests vary, so does the probability of false negatives. This has some repercussions for the application of statistical standards in the interpretation of results. A reader who seeks to assess the impacts of a program variant relative to a pure control group should carefully consider the precision of the estimate. Meanwhile, a reader who directly compares two program variants in the spirit of an A/B test may be less concerned with the thresholds of detectability: if the interventions are cost-equivalent and impacts are a priori expected to be equal, even imprecisely estimated effects can have some epistemic value.

Within each region, three equally sized cohorts of 23 villages each were formed, resulting in six region-cohorts. As displayed in Fig. 1, the randomization was stratified by region-cohort and assigned villages at random to one of five arms, labeled A-E. Eligible participants within each village were further randomly allocated to sub-arms.

In 36 A-type villages, 30 households were assigned to controls (sub-arm A1) and 35 to the microenterprise program (A2). A further 5 households were assigned to a training module designated ex ante to be used for operational research purposes only. In 24 B-type villages, 30 households were assigned to controls (B1) and 35 to a variant of the microenterprise program excluding the savings group components (B2). Here too, a further 5 households were assigned to operational research. In 6 C-type villages, 30 households were assigned to controls (C1) and 35 to a variant of the microenterprise program called business-in-a-box that Village Enterprise opted to evaluate for operational research purposes (C2). In 36 D-type villages, 14 households were assigned to within-village controls (D1); 7 were to plain cash transfers (D2); and 7 were to behaviorally designed cash transfers (D3). In 36 E-type villages, 30 households were assigned to controls (E1). Fig. 2 displays the geographic distribution of villages by arm and region.

Following the randomization, a baseline survey team was provided with a list of intended study invitees. Neither enumerators nor invited respondents were acquainted with the intended treatment assignment, so the decisions to accept the invitation and participate in the research study were independent of the randomization. Participants who opted to participate in the baseline survey were formally recruited into the study.

4.3. Intervention design

The standard microenterprise program (sub-arm A2) was the routine program of Village Enterprise, comprising training, transfers, and mentorship. The training component involved sixteen weekly sessions, each of which took one to three hours (excluding travel time). Of these, the first was an introduction to the program; another session involved the formation of microenterprises; six dealt with savings and the formation, functioning, and governance of savings groups; seven with microenterprise administration; and one with environmental conservation. The total duration of the training was approximately 4 months. Several training sessions into the program, a nominal lump sum cash transfer of 240k

[3] At the time of its administration, the PPI was defined as the Progress-out-of-Poverty Index.

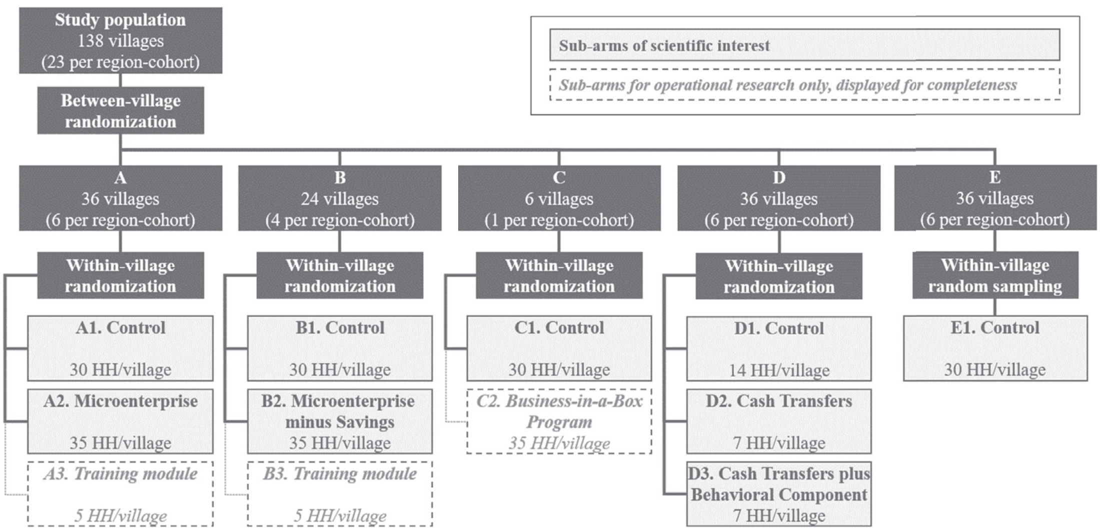

Fig. 1. Assignment of Villages to Study Arms (and Households to Sub-Arms). Note: HH/village denotes the number of households per village.

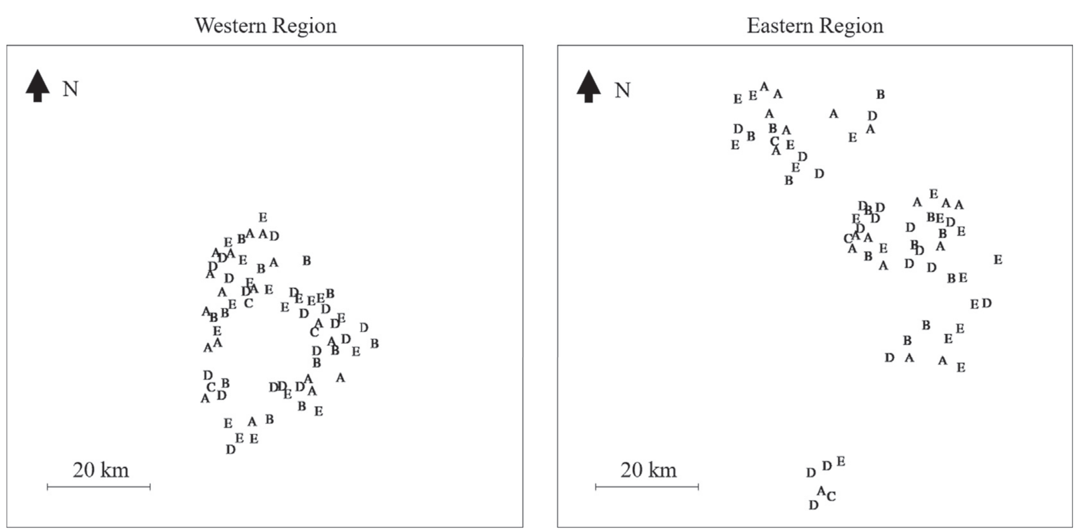

Fig. 2. Maps by Region and Arm. Note: Each letter corresponds to a village; the labels (A-E) define the study arm that was randomly assigned to each village. The position of villages describes its spatial location relative to other villages. Other geographic features are omitted.

Ugandan Shillings (UGX) was made to each business (amounting to UGX 80k per household), contingent upon approval of a business plan. The second transfer (at half the initial amount) was made upon a progress report approximately seven months later, contingent on a review that investments of the initial seed capital had been invested in business activities and that the group was still operating. The average transfer date, weighted by the transfer amounts, was August 2014 (i.e., 15 months before the first and 27 months before the second follow-up survey). Mentorship visits started after the first transfer and continued at monthly intervals.

Sub-arm B2 was a variant of the microenterprise program that excluded the six training sessions on savings group formation, as well as associated coaching visits. Village-level groups with a representative were still formed for the purpose of establishing an administrative counterpart for the implementer.

Sub-arm C2 was a variant of the training program involving the delivery of a pre-selected (typically livestock) asset instead of cash transfers, along with some training on the management of the asset. As discussed above, this arm was excluded from the scientific evaluation at the outset and used only for operational purposes; we discuss it here because the control group in C-type villages (i.e., sub-arm C1) can prove useful in the analysis, and because the activities in these villages will need to be considered in the costing exercise. (Operational research was also conducted in sub-arms A3 and B3; the incremental cost of delivery was however negligible, so these sub-arms will not be revisited in the costing section.)

Sub-arm D2 involved only unconditional cash transfers. Unlike the microenterprise program variants, payments were provided not to three-member businesses, but to households directly. Eligible ones were presented with a voucher and given a time and date when they could expect initial cash disbursements. Intervention leaders explained that a nonprofit had decided to disburse cash for people in the region that they could use as they pleased. The cash disbursement was made in a central village location, with an initial lump sum transfer of UGX 208k per

Table 2
Activity-based costing of sub-arms.

Activity[a]	*Field h/ activity*	Controls: A1, B1, C1, D1, E1	Microenterprise: A2	Microenterprise minus Savings: B2	Business-in-a-Box: C2	Transfers: D2	Transfers plus Behavioral Intervention: D3
Community mapping	*2*	●	●	●	●	●	●
Targeting	*24*	●	●	●	●	●	●
Cash transfer delivery	*12*	○	●	●	●	●	●
Training: business administration	*34*	○	●	●	●	○	○
Training: savings groups	*26*	○	●	○	●	○	○
Training: behavioral	*12*	○	○	○	○	○	●
Training: asset management	*6*	○	○	○	●	○	○
Training: other	*8*	○	●	●	●	○	○
Business group coaching	*60*	○	●	●	●	○	○
Savings group coaching	*6*	○	●	○	●	○	○
Field hours (per savings group)		26	172	140	178	38	50
Trial households		3,324	1,179	791	186	243	237
Total field-hours		2,881	6,760	3,691	1,104	308	395
Cost allocation key[b]		19.03%	44.65%	24.38%	7.29%	2.03%	2.61%
Village Enterprise expenses, USD[c]	**Total**						
Field delivery costs[d]	94,738	18,029	42,303	23,101	6,907	1,926	2,472
Cash transfers[e]	156,326	–	54,145	36,326	8,542	29,015	28,299
Subtotal	*251,064*	*18,029*	*96,448*	*59,427*	*15,448*	*30,941*	*30,771*
Other Ugandan program costs[f]	227,948	43,379	101,785	55,584	16,618	4,635	5,948
Subtotal	*479,012*	*61,407*	*198,233*	*115,011*	*32,066*	*35,576*	*36,718*
Int't program & overhead costs[g]	169,840	32,321	75,838	41,414	12,382	3,453	4,432
Grand total	**648,852**	**93,728**	**274,071**	**156,425**	**44,448**	**39,029**	**41,150**
Cost per household, UGX[h]							
Field delivery costs		14,172	93,756	76,313	97,026	20,714	27,255
Cash transfers		–	120,000	120,000	120,000	312,000	312,000
Subtotal		*14,172*	*213,756*	*196,313*	*217,026*	*332,714*	*339,255*
Other Ugandan program costs		34,100	225,585	183,616	233,454	49,839	65,577
Subtotal		*48,272*	*439,341*	*379,928*	*450,480*	*382,552*	*404,832*
Int'l program & overhead costs		25,407	168,079	136,809	173,943	37,134	48,860
Grand total		**73,680**	**607,420**	**516,737**	**624,423**	**419,686**	**453,692**

Notes:
[a] Field hours by activity are quantified by savings group (the typical unit of training) and include field transport time. Symbol ● indicates that the activity applies to the sub-arm in question. Group C2 is included to enable a full accounting of organizational costs.
[b] We divide the number of field-hours per activity by 30 (i.e., the average savings group size) and multiply it by the number of trial participants to arrive at total field-hours spent per intervention. The cost allocation key is the proportion of total field hours.
[c] Totals are based on internal financial reports of Village Enterprise.
[d] Includes direct compensation and logistical costs associated with field coordinators, trainers, coaches; costed using allocation key.
[e] Costed using exchange rates at intervention time; excludes rate gains/losses from mismatch between withdrawal and disbursement.
[f] Includes internal monitoring & evaluation, administrative, and managerial costs incurred in Uganda. Costed using allocation key.
[g] Includes US-based administrative, managerial, and fundraising costs. Costed using allocation key.
[h] Translated to UGX using the nominal exchange rate at the time of intervention and divided by the number of households in the corresponding sub-arm.

household, followed by a second transfer at half the initial amount. The timing of the two payments mirrored that of the microenterprise program variant. The amounts were budgeted in the planning stage to equate 92.6% of the total expected cost of the microenterprise program. In other words, 7.4% were subtracted to account for targeting, delivery logistics, and overhead; this was the lowest rate of non-cash costs that was identified in a review of cash transfer delivery initiatives.

Sub-arm D3 expanded upon the cash transfers described in sub-arm D2 using a light-touch behavioral intervention that attempted to distill, with the help of a psychologist, the existing training and mentorship activities into their presumed psychological elements. The intervention was comprised of three sessions, including (a) an introductory discussion alongside the provision of the cash vouchers (35 min); (b) a workshop surrounding the first cash disbursement (145 min); and a meeting surrounding the second disbursement (30 min). Goal setting and plan-making activities were designed on the basis of literature on mental contrasting and implementation intentions (Gollwitzer, 1999; Oettingen, 2000). Participants also completed self-affirmation exercises intended to address the stigma of poverty and to promote a sense of adequacy and pride (Cohen and Sherman, 2014; Hall et al., 2014) Participants were asked to think about peers who had been successful and about ways they could follow their peers' examples. This was motivated by work on role models (Lockwood and Kunda, 1997), on social norms (Cialdini and Trost, 1998), and on social comparison processes (Festinger, 1957). Participants also completed drawings and created slogans to help remind them of their goals (Karlan et al., 2010; Rogers and Milkman, 2016). Finally, the program included a mental accounting exercise (Thaler, 1999). The first transfer was provided in two envelopes, with one (amounting to UGX 188k) labeled as intended to support the goal and the other (UGX 20k) as intended for personal incidentals. This was meant to encourage participants to distinguish between consumption and investment.

4.4. Intervention costing

Though the project aimed to align the costs of the cash and microenterprise variants, incurred costs differed somewhat from budgeted ones. In particular, managerial effort turned out higher than anticipated because of the administrative demands associated with the randomized implementation. As we will use incurred costs as the basis of any assessments of returns, Table 2 allocates these (as quantified in the implementer's internal financial reports) to each of the intervention sub-arms; for expenses that cannot be directly allocated, the relative time intensity of activities of the different sub-arms is used as an allocation key ("activity-based costing").

4.5. Data collection

As displayed in Fig. 3, the study builds on three household surveys: one baseline and two follow-ups (labeled midline and endline in the data set). The baseline survey and program implementation were staggered by cohort, while the two follow-up surveys were conducted in the same time period for all study participants.

At the outset of the study, the outcome variables perceived as most central to the theory of change were key poverty indicators (i.e., per-capita consumption, income, and assets). The structure of financial positions (i.e., savings and debt), the employment status of household members, and the subjective well-being of the respondent were also of elevated interest. Further measures on nutrition, education, health, decision-making, cognitive performance, and community life were also collected.

Over the course of the evaluation, some measurement decisions were updated. Diverse psychological and community related measures (e.g., self-control, pride, aspirations, expectations, trust, intimate partner violence) were added to the follow-up surveys. In these follow-up surveys, income and asset measures were collected in updated manner (specifically, collected separately for households and businesses, whereas previously they had been pooled). The collection of cognitive performance data was piloted but abandoned, and no cognitive outcomes were analyzed. Survey forms, data sets, and code are publicly archived in de-identified form.

4.6. Empirical strategy

We start the analysis process by creating an entire universe of plausible analyses. For this purpose, we first identify central dimensions along which different sensible analyses can differ. We distinguish three classes of "choice dimensions": tests, models, and operationalizations. Tests are defined as alternative combinations of outcomes, survey rounds, and intent-to-treat assignments. Each test has a different substantive interpretation. Alternative models and operationalizations do not have a different substantive interpretation, though their choice can influence the results of any given test.

The dimensions along which "tests" can differ include the following:

(1) **Definition of outcomes.** We initially restrict the specification selection exercise to the key poverty indicators. Outcomes y are therefore defined as the consumption, asset, and cash inflow aggregates (as shown in Table 1), as well as the three sub-composites of each. (Total: 12 choices)

(2) **Definition of outcome rounds.** Each poverty outcome can be defined in terms of follow-up round F: the first follow-up (midline); the second follow-up (endline); and a pooled value averaging the two. (Total: 3 choices)

(3) **Definition of comparisons.** In defining intent-to-treat assignment T, we seek to estimate 6 types of impacts:

[a] those of spillovers by comparing the set of sub-arms A1, B1, C1, and D1 to E1 (which may help in the selection of appropriate counterfactuals where alternative choices are available);

[b] those of the microenterprise program variants by comparing the set of sub-arms A2 and B2 to untreated controls (though alternative choices of appropriate control counterfactuals are available, as described below);

[c] those of the cash transfer program variants by comparing the set of sub-arms D2 and D3 to untreated controls (though alternative choices of appropriate control counterfactuals are available, described below);

[d] those of the savings group component, conditional on the microenterprise program, by comparing sub-arm A2 to B2;

[e] those of the behavioral intervention component, conditional on the cash transfer program, by comparing sub-arm D3 to D2; and

[f] the incremental impacts of the microenterprise program over those of the cash transfer program by comparing the set of sub-arms A2 and B2 to the set of sub-arms D2 and D3.

We define "models" as alternative forms of specifying the econometric analysis for any given test. Choice dimensions include the following:

(1) **Use of baseline values.** We can control for the baseline value of the outcome y_B; subtract it from outcome data for differences in differences; and leave it out of the estimation process altogether. (Total: 3 choices)

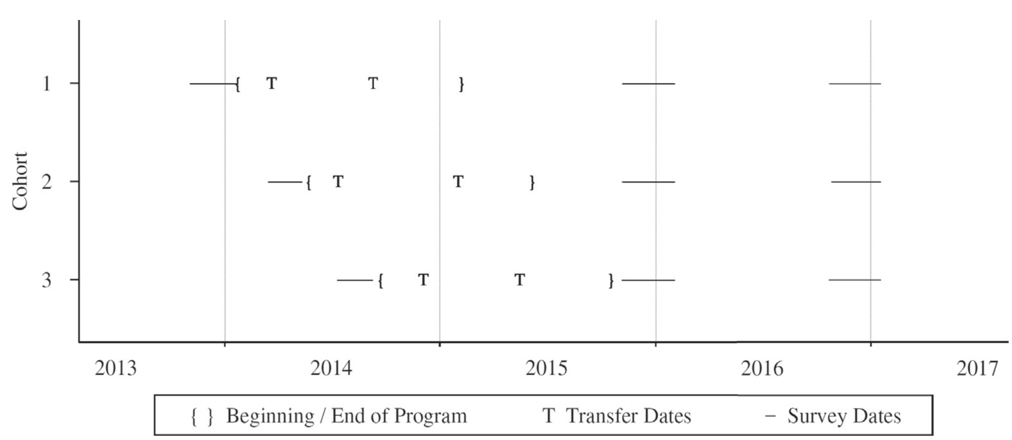

Fig. 3. Intervention and survey timelines by study cohort.

Table 3
Covariate balance by comparison set.

	Comparison set: Average value	[a]	[b][i]	[b][ii]	[b][iii]	[c][i]	[c][ii]	[c][iii]	[d]	[e]	[f]
		p values									
HH size	5.91	0.21	0.62	0.38	0.85	0.26	0.32	0.67	0.66	0.50	0.71
Age of HH Head	43.1	0.58	0.19	0.46	0.08	0.54	0.13	0.16	0.58	0.94	0.04
HH Head's years of schooling	5.32	0.97	0.44	0.98	0.81	0.56	0.97	0.91	0.37	0.25	0.99
HH Head is female	0.29	0.67	0.44	0.75	0.97	0.63	0.96	0.76	0.59	0.67	0.70
HH Head is monogamously married	0.56	0.77	0.88	0.97	0.80	0.14	0.13	0.07	0.41	0.24	0.09
HH Head is literate	0.47	0.81	0.25	0.81	0.95	0.31	0.74	0.86	0.91	0.28	0.98
HH has iron roof	0.26	0.59	0.02	0.99	0.47	0.62	0.98	0.60	0.86	0.76	0.91
HH has mud walls	0.40	0.99	0.90	1.00	0.98	0.60	0.88	0.84	0.80	0.90	0.89
HH has earth floor	0.97	0.16	0.53	0.35	0.91	0.48	0.11	0.26	0.41	0.64	0.31
HH has sanitary toilet/latrine	0.41	0.54	0.33	0.72	0.97	0.34	0.60	0.78	0.34	0.21	0.87
HH uses wood as main cooking fuel	0.98	0.90	0.03	0.13	0.02	0.51	0.47	0.32	0.54	0.41	0.62
HH uses electric light	0.02	0.64	0.55	0.97	0.72	0.22	0.14	0.19	0.26	0.27	0.16
HH owns its home	0.88	0.61	0.29	0.97	0.61	0.87	0.88	0.85	0.07	0.30	0.92
All HH members have two sets of clothes	0.62	0.87	0.18	0.46	0.27	0.11	0.39	0.29	0.39	0.62	0.13
All HH members have a pair of shoes	0.23	0.36	0.96	0.29	0.38	0.24	0.53	0.17	0.97	0.93	0.12
Test of joint orthogonality		**0.98**	**0.29**	**0.97**	**0.57**	**0.76**	**0.66**	**0.61**	**0.86**	**0.89**	**0.42**

Notes: Data are derived from the baseline. The first column lists the covariates. The first three are continuous variables; and the others are binary. The second column shows the average value of the covariate across all observations. In the case of binary covariates, this represents the proportion of households for which the dummy is coded to 1 ("yes"). The subsequent columns are derived from nine linear regressions (one per comparison set) of the form $T_{ij} = \alpha + \beta X_{ij} + \varepsilon_{ij}$, where T is the applicable intent-to-treat assignment dummy of household i in village j. The p values correspond to each coefficient in the vector of covariates X. Standard errors are adjusted for cluster robustness in so-called clustered comparison sets, i.e.: [a], [b][ii], [b][iii], [c][ii], [c][iii], [d], and [f]. The test of joint orthogonality reports the p value of the F statistic for each of the nine regressions.

(2) **Use of socioeconomic covariates.** Many different socioeconomic baseline covariates X_B, and many more combinations, are available. We can reduce this complexity by pursuing 2 choices: 'selecting none' or 'selecting some set'. When "selecting some set", we use least angle regression (Efron et al., 2004) to identify five among those candidates listed in Table 3.

(3) **Use of fixed effects.** The model can use intercept α or control for α_j, where j is the village cluster; the latter implies the use of cluster fixed effects. (Total: 2 choices)

By "operationalization", we mean the coding of variables. We define the following choice dimensions:

(1) **Outlier adjustment.** As poverty measures are sensitive to outliers, some adjustment is required. To avoid introducing attenuation bias, it is most sensible to adjust each combination of y and T separately. But the appropriate level is not certain ex ante. We recode the highest and lowest 0.5%, 2.5%, or 5% of observations to the cutoff value; i.e., winsorize at the 99%, 95%, or 90% level. (Total: 3 choices)

(2) **Valuation approach.** Where the computation of y involves calculating the value of goods, we use the price estimates reported by respondents; the median prices in a regional geographic unit; and a combination that uses the former where available and the latter where respondents are unsure. (Total: 3 choices)

(3) **Definition of the counterfactual in controlled comparisons.** As defined above, comparison sets [b] and [c] compare a treatment group with a counterfactual group comprised of "pure controls" – i.e., one where nobody is treated. There are different ways to operationalize these controls: we code treatment assignment T to the value zero [i] for controls within villages (within-village controls); [ii] for controls in pure control villages (between-village controls); or [iii] for all available controls ranging from A1 to E1. These choices come with different merits: electing between-village controls circumvents adjustments for cluster robustness, with benefits for statistical power, while selecting only control villages would minimize susceptibility to bias emerging from within-village spillovers. The third option is a compromise between power and unbiasedness. (Total: 3 choices)

Multiplying each of the 12 × 3 × 6 tests with 2 × 3 × 2 alternative models and 3 × 3 × 3 alternative operationalizations would yield a total of 69,984 impact coefficients. But not every specification is applicable for every test. First, a choice of three alternative counterfactuals is only available for comparison sets [b] and [c], but not for comparison sets [a], [d], [e], and [f]. Second, a choice of cluster fixed effects is only available for comparisons within arms, where the unit of randomization as well as the unit of observation is the household (we label "non-clustered comparisons"), because cluster fixed effects would be collinear with the unit of randomization this is itself the cluster (we label these "clustered comparisons"). Third, the use of any valuation other than the respondent's is only appropriate for measures with commodity character. This leaves the number of actual estimates at 16,848, i.e., an average of 78 specifications for each of the 216 tests on average.

For the purpose of operationalizing variables, we rely on so-called specification curves (Simonsohn et al., 2015). These visually plot, for each test, all results associated with all combinations of plausible specifications, allowing us to visually analyze – and illustrate to the reader – how sensitive the results are to alternative operationalization choices.

Once operationalization choices have been made,[4] we engage in model selection. Algorithmic tools can analyze which models best are best supported by the data and identify those that optimize the trade-off between fit and parsimony (Burnham and Anderson, 2004, 2010; Clyde, 2003; Hoeting, 2002; Hoeting et al., 1999). So-called Akaike weights provide proportional support for a given model within a set of models K. To calculate these, we build on the insight that the Akaike Information Criterion AIC is an unbiased estimator of twice minus the log likelihood of a model, and that $e^{-0.5AIC}$ estimates the likelihood of a model conditional on the data (Burnham and Anderson, 2004; Wagenmakers and Farrell, 2004). The proportional support for model x within set K is therefore defined as

$$w_x = \frac{e^{-0.5AIC_x}}{\sum_{k=1}^{K} e^{-0.5AIC_k}} \tag{I}$$

[4] A previous version of this paper conducted variable operationalization and model selection in the reverse order. The change was motivated in the peer review process.

We will define K as the set of models available for each test.[5] We will repeat this procedure for all 216 tests, then select – separately for each test – the customized model with the strongest support. A Bayesian interpretation is that model weights are posterior probabilities conditional on the data, assuming that prior probabilities had been equally distributed across all models. The model with the highest weight is most likely to be the best model, given the available data.

We will be left with 216 preferred estimates: 36 intent-to-treat coefficients and associated p values (i.e., one for each of the 12 key poverty outcomes and three survey rounds) across six comparison groups. To account for multiple inference among these residual estimates, we control for the false discovery rate (Benjamini and Hochberg, 1995) and report sharpened q values following the two-stage method coded described by Anderson (2008). We apply these adjustments across all estimates within a given comparison group and outcome class; put differently, we apply them separately to each set of p values displayed in each individual table in the Online Appendix of Tables. Each of these tables corresponds to our definition of the individual hypothesis.

5. Results

5.1. Balance checks

As discussed, we are evaluating six comparisons; for two of them (i.e., the ones for which the counterfactual is an untreated control group), three sets of counterfactuals are available. This initially brings us to nine available comparison sets; for ease of reference, they are illustrated in Fig. 4.

We first test for covariate balance in each of these comparison sets. As illustrated in Table 3, all comparison sets are well balanced at baseline.

As some indications of differential attrition are discernible among the sub-arms in Table 12 (see "Appendix C: Participant Flow"), we test for differential attrition in each of the comparison groups. The results are presented in Table 4. Comparison sets [c] and [f] are indeed afflicted by differential attrition in both the mid- and endline. We assess the robustness of these comparisons to alternative assumptions about attritors, following the trimming procedures proposed by Lee (2018) to put bounds on the treatment effects. We limit this procedure to poverty outcomes.

5.2. Operationalization

Specification curves plot test statistics as a function of a wide range of plausible specification choices. While they can directly assist statistical inference (Simonsohn et al., 2015), they can also be used to merely complement other analyses. We will use them to inform the operationalization of our data. To illustrate the analysis and decision process, Fig. 5 plots alternative impact estimates of the cash transfer program variants – i.e., comparison [c] – on consumption. We can see that estimates are not heavily correlated with the choice of the outlier correction nor the valuation rule, while the choice of the counterfactual is highly influential: estimates are much more pronounced in within-village comparisons. This indicates negative within-village spillovers; but as between-village controls and all available controls seem to yield comparable results, it appears that these spillovers are specific to cash transfer villages. Indeed, no specification curve investigating comparison set [a] indicates significant spillovers in the aggregate (see Online Appendix of Specification Curves).

For reference, the specification curve showing the impacts of the microenterprise program is displayed in Fig. 8, and all others are in the Online Appendix of Specification Curves. Based on these, we opt to winsorize all outcomes at the 95% level and to value all commodities using local median prices. As there is no evidence of spillovers in the aggregate, we opt to use all available controls as our preferred specification. A section on robustness checks will investigate the potential repercussions of bias from spillovers that are specific to cash villages.

5.3. Model selection

We retain the estimates associated with the chosen operationalizations and proceed to the model selection process. We start by extracting AIC for each of the remaining estimates and calculating Akaike weight w for each. This quantifies a conditional probability between zero and one for each model in each test; a higher weight w implies a better fit. By identifying the highest w within each set K, we can identify a customized model for each test. Fig. 6 shows the conditional probabilities of the best available model by test. Where w approaches 1, the algorithm is unambiguous in its recommendation.

Fig. 6 also shows the probabilities associated with a "universal" model – that with the strongest support *across* tests. It becomes apparent that our inference is virtually identical if we restrict ourselves to this single model. It is the variant that controls for baseline values as well as socioeconomic characteristics, and that applies fixed effects where possible. More formally, it is defined as:

$$y_{ijF} = \alpha_j + \beta T_{ij} + \gamma y_{ijB} + \delta X_{ijB} + \varepsilon_{ij} \tag{III}$$

Here, y_{ijF} is the per-capita outcome in household i in village j at the time of follow-up F; T is the randomized assignment, coded to 1 for intent-to-treat and to 0 for the counterfactual; y_{ijB} is the baseline observation of the outcome; and X_{ijB} is a set of socioeconomic baseline covariates. The coefficient for the intent-to-treat estimate is β. α_j denotes village fixed effects; where these are not applicable (i.e., in so-called clustered comparisons), the intercept term is limited to constant α.

Given that a single model is very close to optimal, we can achieve gains in clarity by reporting all results using this model. This allows us to summarize all specification choices in Fig. 7. A section on robustness checks will showcase the minor differences that emerge when applying the customized model in those two cases where it differs from the universal one.

For all outcomes other than the set of main economic poverty indicators, we apply the same universal model; these outcomes do not feed back into the model selection process. Where outcome data are available at the individual level, our unit of analysis will correspond to the individual as opposed to the household. Logistic regression is applied whenever outcomes are binary, and estimates presented as odds ratios. As in the core poverty outcomes, we apply false discovery adjustments for each comparison group and outcome group separately, so each table presented in the Online Appendix of Tables is subject to a dedicated false discovery rate adjustment.

5.4. Impact estimates

Table 5 shows the impact estimates on the main economic aggregates and Table 6 shows them by sub-aggregate. Composite indices of non-economic outcomes are summarized in Table 7.

Numerous positive impacts can be identified in the microenterprise program variant, as defined in the comparison set [b][iii]. Impacts on consumption (Table 5) are driven predominantly by gains in food and beverage consumption (Table 6), which is corroborated by nutritional impacts (Table 7). Half of the consumption gains are explained by gains in cash income. Gains in asset stock are driven by gains in livestock

[5] Both numerator and denominator can take on explosive values, making (I) impossible to compute (Burnham and Anderson, 2004). In practice, we therefore compute Akaike weights as

$$w_x = \frac{e^{-0.5(AIC_x - AIC_{\min})}}{\sum_{k=1}^{K} e^{-0.5(AIC_k - AIC_{\min})}} \tag{II}$$

Here $AIC_{\min}$ is defined as the minimum Akaike Information Criterion in set K.

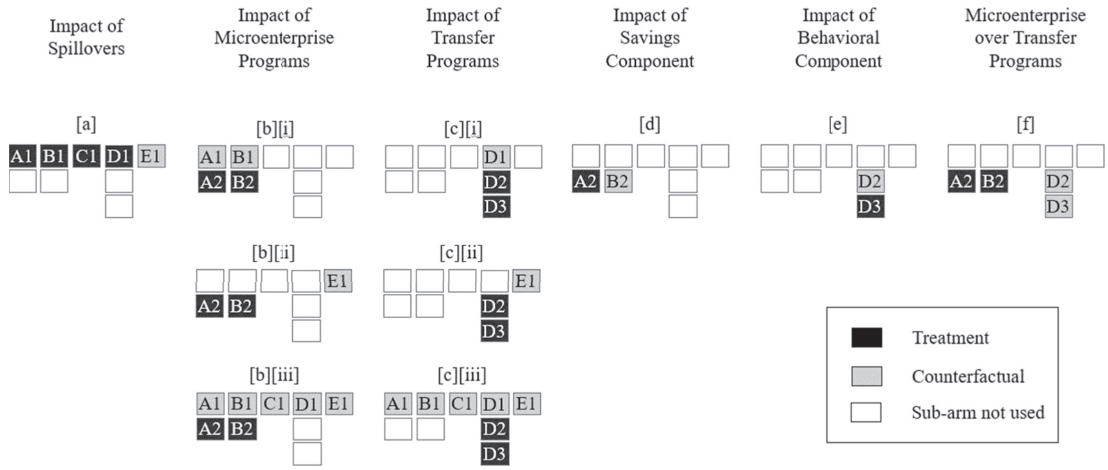

Fig. 4. Available Comparison Sets. Notes: For the definition of sub-arms A1, A2, etc, consult Fig. 1. For more detailed definitions of comparisons [a]-[f] and alternative counterfactuals [i]-[iii], consult the empirical strategy section.

Table 4
Tests of differential attrition by comparison set.

Comparison set		Midline							Endline						
		Treatment			Counterfactual				Treatment			Counterfactual			
		Surveyed	Attrited	Odds	Surveyed	Attrited	Odds	p value	Surveyed	Attrited	Odds	Surveyed	Attrited	Odds	p value
[a]		2,164	165	0.076	933	62	0.066	0.530	2,113	216	0.102	915	80	0.087	0.386
[b]	[i]	1,819	151	0.083	1,584	115	0.073	0.297	1,783	187	0.105	1,543	156	0.101	0.747
[b]	[ii]	1,819	151	0.083	933	62	0.066	0.322	1,783	187	0.105	915	80	0.087	0.332
[b]	[iii]	1,819	151	0.083	3,097	227	0.073	0.348	1,783	187	0.105	3,028	296	0.098	0.530
[c]	[i]	463	17	0.037	430	33	0.077	0.016 **	453	27	0.060	419	44	0.105	0.026 **
[c]	[ii]	463	17	0.037	933	62	0.066	0.092 *	453	27	0.060	915	80	0.087	0.227
[c]	[iii]	463	17	0.037	3,097	227	0.073	0.020 **	453	27	0.060	3,028	296	0.098	0.076 *
[d]		1,073	106	0.099	746	45	0.060	0.027 **	1,059	120	0.113	724	67	0.093	0.340
[e]		229	8	0.035	234	9	0.038	0.846	223	14	0.063	230	13	0.057	0.791
[f]		1,819	151	0.083	463	17	0.037	0.010 **	1,783	187	0.105	453	27	0.060	0.057 *

Notes: p values are derived from logistic regression without covariates. Standard errors are adjusted for cluster robustness in so-called clustered comparisons, i,e: [a], [b][ii], [b][iii], [c][ii], [c][iii], [d], and [f]. For the definition of comparison sets, consult Fig. 4.

ownership. While impacts on financial assets are detectable, they are low in absolute terms. Alongside improvements in living conditions, microenterprise participants report gains in psychological outlook and the social conditions in their communities. Given the household size from Table 1 and the program costs from Table 2, the payback period – i.e., the number of years that it takes for consumption impacts to recoup the initial investment – is roughly for years. That said, it is not clear if effects do in fact persist. Table 13 disaggregates effects by survey round; though the two estimated impacts of the microenterprise program lie within each other's confidence intervals, point estimates appear consistent with a possible diminution of effects over time. We will explore the repercussions in the robustness checks.

Savings groups do not appear to enhance asset accumulation: we fail to detect an impact even on the narrow definition of financial assets.[6] Savings groups do however seem to increase self-employment activities, and there are indications that perceptions of social conditions in the community improved.

The light-touch psychological intervention also appeared to manipulate psychological outlook, and it did tend to lead to higher asset stocks. But it did not result in improvements in the standard of living as indicated by the economic flow variables (i.e., cash inflows and consumption): combining this intervention with cash transfers did not appear to effectively substitute the integrated graduation-style approach. The results do not suggest that the impacts of the integrated microenterprise program are mediated exclusively via changes in psychological outlook.

Few positive impacts are discernible from the pure cash transfer program.[7] Recipients do report lower loan balances[8] and net assets impacts are comparable to those of the microenterprise program. But consider that the incremental asset stock must have been substantially greater at the time of program initiation – akin to UGX 35k per capita on average in the cash transfer program, compared to UGX 20k in the case of the microenterprise program; this suggests assets depreciated or were dissaved at higher rates in the cash transfer arms. Yet surprisingly, no effects are detectable on short-term consumption flows; short-term effects tend negative. Unlike in the microenterprise program, no encouraging signals emerge on psychological, nutritional, or social outcomes. This is not explained by a reduction in labor force participation: in fact, Table 16 suggests that labor inputs increased.

6 An alternative approach to measuring savings positions might involve consulting administrative data on balances in the savings groups established by Village Enterprise. We do not use these data, as they are only available for the sub-arm A2 where this activity was conducted. However, it should be noted that these yield significantly higher positions than self-reported ones provided by survey respondents, pointing to possible under-reporting in the survey.

7 School attendance and enrolment do tend to increase; see Table 14.

8 See Table 15.

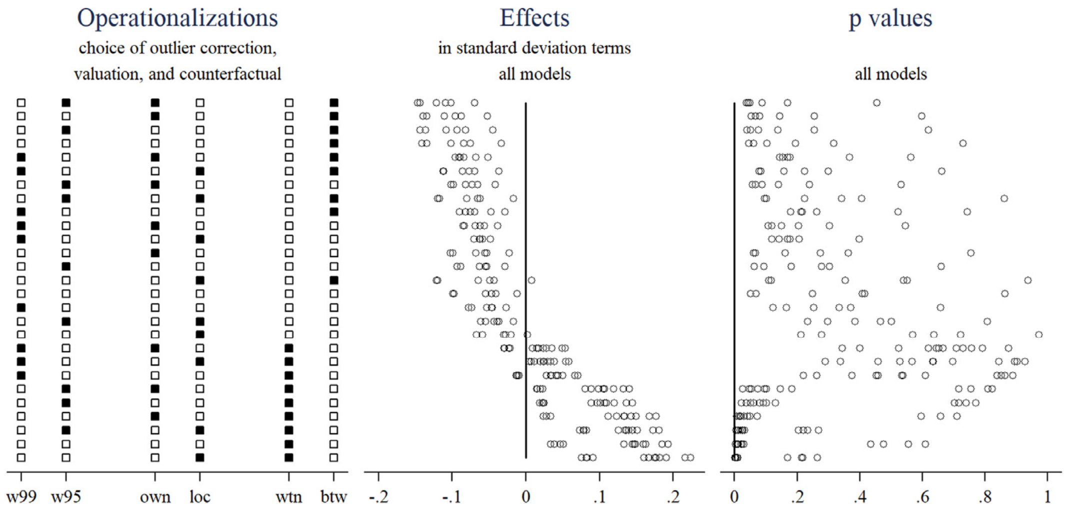

Fig. 5. Impact of Transfer Programs on Consumption (Specification Curve). Notes: Each row in the chart on the left describes one combination of operationalization choices. In the same row, the charts on the right show the corresponding effect sizes and p values for all available models. More specifically, read the three charts as follows:

- "**Specification alternatives**": Columns define specification features. A black square ■ indicates that a choice dimension applies, and a blank square □ that it doesn't. Where two columns are adjacent, three alternatives are available; the third column is not displayed, as it can be inferred that it applies whenever the other two do not apply. The first two columns define the choice dimension of outlier adjustment. w99 implies that 0.5% of highest and 0.5% lowest per capita outcomes are recoded to the cutoff value, and w95 implies that 2.5% of highest and 2.5% lowest per capita outcomes are recoded to the cutoff value. Where symbols in both columns are blank, a third choice (90% winsorization) is applied. The next two columns define the valuation approach that is used. Own implies that only the respondent's valuation is used; loc implies that regional prices (specific to the survey round) are used. Where symbols in both columns are blank, a third option is applied that uses own values except where these are unavailable, in which case loc values are used. (Some goods, such as jewelry assets, are too heterogeneous to allow for a sensible unit valuation across households; for such categories, only the respondent's own valuation is always used. When aggregated with measures that use another valuation rule, the latter valuation rule is displayed.) The final two columns define the counterfactual. Alternative choices are only applicable in comparison sets [b] ("Impact of Microenterprise Programs") and [c] ("Impact of Transfer Programs). **wtn** implies a comparison within villages, and **btw** implies a between-village comparison. Where symbols in both columns are blank, a third choice applies, and all control groups (so sub-arms A1, B1, C1, D1, and E1) are used as the counterfactual. Note that the paper refers to the first choice as a clustered comparison, and to the latter two as non-clustered comparisons. Following Abadie et al. (2017), standard errors are clustered in clustered comparison sets.
- **"Effect sizes":** This displays the intent-to-treat estimates that correspond to the specification choice, presented in terms of standard deviations of the counterfactual. Each hollow dot represents one model. Operationalizations are sorted in ascending order of median treatment effects for visual ease.
- **"p values":** This displays the p values corresponding to estimates, again showing one dot per model.

6. Robustness checks

6.1. Sensitivity to model selection

Given that the algorithm recommended the same model for over 99% of tests, we restricted ourselves to this model throughout the paper to simplify the interpretation of results. Table 8 loosens this restriction and shows the results that deviate from the universal model when we take advantage of the algorithm's ability to customize a model to each tests. Two of the 216 estimates – both associated with comparison [e], so estimating the impact of the behavioral component – are very slightly attenuated.

6.2. Sensitivity to assumptions about attriters

Table 4 found that study participants in the cash transfer groups attrited at significantly lower rates than ones in the pure control (comparison set [c]) and the microenterprise groups (comparison set [f]). Table 9 uses the procedure by Lee (2018) to put bounds on the estimates found in these comparison sets. The impact estimates are not robust to this: no finding is significant under both conservative and aggressive assumptions about the nature of differential attrition. Put differently, the measured relative and absolute impacts of cash transfers are biased if we presume that true impacts are correlated with respondents' propensity to respond to the follow-up surveys. While this concern exists even in the presence of minimal attrition, it is expanded when attrition is significant.

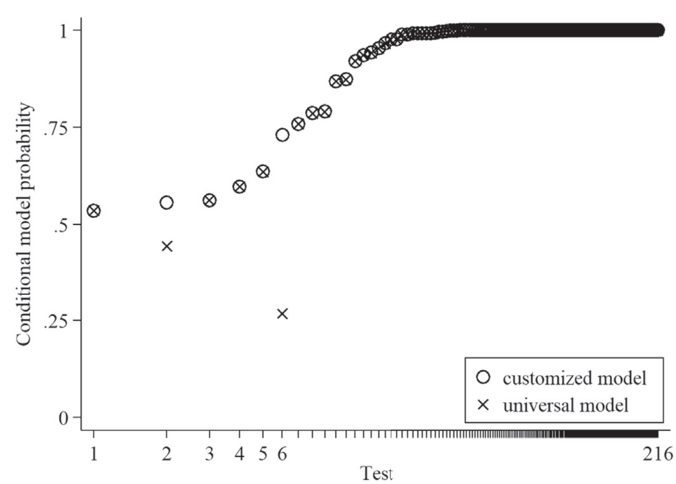

Fig. 6. Conditional Probabilities of High-Quality Models, by Test. Note: The conditional probabilities of alternative models are shown for each of the 216 tests. The customized model is that with the highest conditional probability in each test, while the universal model is the single model with the highest conditional probability across tests. We can see that the universal model is the best available one in 214 of the 216 tests. For visual ease, tests are ordered by the probability of the customized model; conditional probabilities of inferior models are omitted; and the scale of the x axis is logarithmic.

			Tests					
			Impact of Spillovers	Impact of Microenterprise Programs	Impact of Transfer Programs	Impact of Savings Component	Impact of Behavioral Component	Impact of Microenterprise over Transfer Programs
		Comparison set	[a]	[b][iii]	[c][iii]	[d]	[e]	[f]
Analytical specification choices	Model selection	Difference-in-differences (did)	□	□	□	□	□	□
		ANCOVA (anc)	■	■	■	■	■	■
		Cluster fixed effects (fe)	□	□	□	□	■	□
		Socioeconomic covariates (cvt)	■	■	■	■	■	■
	Operationalization	95% winsorization (w95)	■	■	■	■	■	■
		99% winsorization (w99)	□	□	□	□	□	□
		Self-reported unit valuation (own)	□	□	□	□	□	□
		Medial local unit valuation (loc)	■	■	■	■	■	■
		Within-village comparison (wtn)	□	□	□	□	■	□
		Between-village comparison (btw)	■	□	□	■	□	■
	Other	Clustered standard errors	■	■	■	■	□	■
		Robustness check: differential attrition	□	□	■	□	□	■

Fig. 7. Summary of Analyses. Notes: Each column describes the specification for the corresponding comparison set. A black square ■ indicates that a choice dimension applies, and a blank square □ that it doesn't. Where two rows are adjacent are displayed, three alternatives are available; the third row is not displayed, as it can be inferred that it applies whenever the other two do not. (E.g., the counterfactual in comparison sets [b][iii] and [c][iii] are neither within-village nor between-village controls, but all available controls.) Following Abadie et al. (2017), standard errors are clustered in clustered comparison sets.

Table 5
Intent-to-Treat Estimates, Economic Aggregates (per capita).

Comparison set	Impact of Spillovers	Impact of Microenterprise Programs	Impact of Transfer Programs	Impact of Savings Component	Impact of Behavioral Component	Impact of Microenterprise over Transfer Programs
	[a]	[b][iii]	[c][iii]	[d]	[e]	[f]
Total Consumption						
Coefficient	−16,462	26,061	−17,141	8,833	−24,982	46,294
Error	18,915	11,248	19,679	21,944	29,279	22,429
Effect/mean	−0.02	0.04	−0.03	0.01	−0.04	0.07
Effect/sd	−0.04	0.07	−0.04	0.02	−0.07	0.13
p value	0.386	0.022 **	0.385	0.689	0.394	0.042 **
q value	1.000	0.055 *	0.580	1.000	1.000	0.193
N	3,094	4,906	3,545	1,812	451	2,263
Total Net Assets						
Coefficient	−3,640	16,343	15,852	−5,917	19,283	−831
Error	6,789	5,449	8,397	9,048	11,479	9,627
Effect/mean	−0.03	0.14	0.13	−0.04	0.15	−0.01
Effect/sd	−0.03	0.12	0.12	−0.04	0.15	−0.01
p value	0.593	0.003 ***	0.061 *	0.516	0.094 *	0.931
q value	1.000	0.021 **	0.132	1.000	0.451	1.000
N	3,004	3,796	2,773	1,746	462	1,819
Total Productive Cash Inflows						
Coefficient	−8,069	13,483	−8,453	20,208	−5,154	12,983
Error	9,273	6,747	11,740	11,007	17,309	11,309
Effect/mean	−0.06	0.10	−0.06	0.15	−0.04	0.09
Effect/sd	−0.04	0.07	−0.04	0.10	−0.03	0.07
p value	0.386	0.048 **	0.473	0.071 *	0.766	0.254
q value	1.000	0.087 *	0.599	0.506	1.000	1.000
N	2,529	4,021	2,916	1,885	472	2,354

Notes: The preferred specification is used. Outcomes are pooled across survey rounds. Coefficients are reported in current Ugandan Shillings (UGX). Effect/mean is the coefficient divided by the mean outcome in the counterfactual. Effect/sd is the coefficient divided by the standard deviation of the outcome in the counterfactual. False discovery rate adjustments that form the basis of q values are calculated on the sets of results as defined in each of the tables in the Online Appendix of Tables. Standard errors are adjusted for cluster robustness in all comparison sets except [e]. For more detailed results, consult Table 6 and the Online Appendix of Tables.

Table 6
Intent-to-Treat Estimates, Economic Sub-Aggregates (per capita).

		Comparison set	Impact of Spillovers	Impact of Microenterprise Programs	Impact of Transfer Programs	Impact of Savings Component	Impact of Behavioral Component	Impact of Microenterprise over Transfer Programs
			[a]	[b][iii]	[c][iii]	[d]	[e]	[f]
Consumption	Food and Beverage	Coefficient	1,169	25,180	−10,261	15,944	−37,416	38,623
		Error	13,429	9,381	15,479	18,221	23,907	18,194
		p value	0.931	0.008 ***	0.508	0.385	0.118	0.036 **
		q value	1.000	0.029 **	0.599	1.000	0.463	0.193
	Recurring	Coefficient	−5,188	−1,402	−7,690	−1,069	−2,127	6,039
		Error	3,420	1,917	2,819	3,282	4,574	2,924
		p value	0.132	0.466	0.007 ***	0.746	0.642	0.042 **
		q value	1.000	0.296	0.045 **	1.000	1.000	0.193
	Infrequent	Coefficient	−5,437	2,839	1,171	−4,145	13,121	442
		Error	4,045	2,605	4,352	4,442	6,407	4,546
		p value	0.181	0.278	0.788	0.355	0.041 **	0.923
		q value	1.000	0.218	0.713	1.000	0.335	1.000
Net Assets	Livestock	Coefficient	−2,671	10,584	15,155	−2,438	19,185	−3,504
		Error	3,846	2,657	4,728	4,900	7,790	5,383
		p value	0.489	0.000 ***	0.002 ***	0.621	0.014 **	0.517
		q value	1.000	0.002 ***	0.032 **	1.000	0.335	1.000
	Durable	Coefficient	917	4,440	2,172	−3,794	163	1,864
		Error	3,007	2,452	4,416	3,830	5,684	4,664
		p value	0.761	0.072 *	0.624	0.326	0.977	0.690
		q value	1.000	0.099 *	0.646	1.000	1.000	1.000
	Net Financial	Coefficient	812	1,238	3,041	996	1,101	−1,749
		Error	687	572	1,059	964	1,661	1,087
		p value	0.240	0.032 **	0.005 ***	0.306	0.508	0.111
		q value	1.000	0.070 *	0.043 **	1.000	1.000	0.527
Productive Cash Inflows	Net Farming Inflows	Coefficient	−5,638	1,409	−2,612	−876	5,863	1,224
		Error	3,508	2,161	5,260	4,188	7,007	5,555
		p value	0.110	0.515	0.620	0.835	0.403	0.826
		q value	1.000	0.304	0.646	1.000	1.000	1.000
	Other Self-Employment	Coefficient	−5,648	11,862	6,417	20,169	−7,411	6,927
		Error	5,235	4,361	6,881	6,792	10,889	6,897
		p value	0.283	0.007 ***	0.353	0.004 ***	0.496	0.318
		q value	1.000	0.029 **	0.575	0.082 *	1.000	1.000
	Paid Employment	Coefficient	800	−1,088	−3,472	−2,198	−11,333	628
		Error	3,529	2,681	4,784	4,005	5,581	4,445
		p value	0.821	0.686	0.469	0.585	0.043 **	0.888
		q value	1.000	0.378	0.599	1.000	0.335	1.000

Notes: The preferred specification is used. Outcomes are pooled across survey rounds. Coefficients are reported in current UGX. For results by survey round, consult the Online Appendix of Tables. False discovery rate adjustments that form the basis of q values are calculated on the sets of results from these tables. Standard errors are adjusted for cluster robustness in all comparison sets except [e].

Table 7
Intent-to-treat estimates, composite indices.

Comparison set	Impact of Spillovers	Impact of Microenterprise Programs	Impact of Transfer Programs	Impact of Savings Component	Impact of Behavioral Component	Impact of Microenterprise over Transfer Programs
	[a]	[b][iii]	[c][iii]	[d]	[e]	[f]
Nutrition						
Coefficient	0.012	0.135	0.021	−0.019	0.066	0.083
Error	0.050	0.034	0.050	0.060	0.079	0.062
p value	0.809	0.000 ***	0.670	0.757	0.399	0.183
q value	1.000	0.001 ***	1.000	1.000	1.000	1.000
Psychological Outlook						
Coefficient	−0.126	0.143	0.107	−0.060	0.178	0.033
Error	0.053	0.042	0.067	0.066	0.105	0.071
p value	0.019 **	0.001 ***	0.117	0.375	0.091 *	0.644
q value	0.264	0.040 **	0.880	1.000	0.819	1.000
Social Conditions						
Coefficient	0.079	0.088	−0.025	0.165	0.022	0.109
Error	0.060	0.041	0.061	0.074	0.112	0.069
p value	0.190	0.032 **	0.681	0.030 **	0.847	0.116
q value	0.939	0.309	1.000	0.258	1.000	0.534

Notes: The preferred specification is used. All outcomes are standardized composite indices, following the method used by Kling et al. (2007). Outcomes are pooled across survey rounds. For results by survey round, consult the Online Appendix of Tables. False discovery rate adjustments that form the basis of q values are calculated on the sets of results from these tables. The nutrition index is the aggregate of the Household Dietary Diversity Score and the inverse of the Household Food Insecurity Access Score. The psychological outlook index is the aggregate of subjective well-being, aspirations, self-control, sense of control, sense of status, and sense of pride. The social conditions index is the aggregate of sense of community, sense of trust, risk sharing, empowerment of women, and the inverse of intimate partner violence. For results by survey round, consult the Online Appendix of Tables. False discovery rate adjustments that form the basis of q values are calculated on the sets of results from these tables. This appendix also shows those outcomes that are not averaged across survey rounds, including all binary outcomes in the domains of labor, schooling, and health.

Table 8
Selected estimates in comparison group [e], customized model.

	Infrequent Consumption, Pooled Follow-ups		Net Financial Position, First Follow-up	
	βcustomized	βcustomized - βuniversal	βcustomized	βcustomized - βuniversal
Coefficient	11,527	−1,594	−745	−3
Error	6,389		1,613	
Effect/mean	0.16		−0.19	
Effect/sd	0.17		−0.03	
p value	0.072 *		0.644	
q value	0.422		1.000	
N	473		462	

Notes: Outcomes are pooled across survey rounds. Coefficients are reported in current UGX, rounded to the nearest shilling. $\beta_{customized}$ are the coefficients associated with those two tests in Fig. 6 in which the customized model outperformed the universal one. In the test of infrequent consumption, the customized model excludes the socioeconomic covariates; in the test of net financial position, it excludes the baseline covariate. $\beta_{universal}$ is the previously reported estimate.

6.3. *Sensitivity to assumptions about spillovers*

We found indications of negative spillovers in cash arm D. While spillovers are not significant in the aggregate, they tend negative. We repeat our calculation under the assumption that spillovers have tainted our within-village control groups. Table 10 illustrates the impact estimates for the microenterprise and cash transfer programs when using only control villages (sub-arm E1) as the counterfactual. Confidence

Table 9
Lee bounds.

		Conservatively Trimmed Estimate			Aggressively Trimmed Estimate		
		Total Consumption	Total Net Assets	Total Productive Cash Inflows	Total Consumption	Total Net Assets	Total Productive Cash Inflows
Impact of Transfer Programs [c][iii]	Coefficient	−48,001	287	−35,716	−6,417	18,420	−992
	Error	17,043	7,044	8,649	20,379	8,516	11,704
	p value	0.006 ***	0.968	0.000 ***	0.753	0.032 **	0.933
Impact of Microenterprise over Transfer Programs [f]	Coefficient	33,190	−4,528	4,526	79,796	16,065	39,236
	Error	23,372	9,798	11,352	19,282	7,126	9,032
	p value	0.159	0.645	0.691	0.000 ***	0.026 **	0.000 ***

Notes: Outcomes are pooled across survey rounds. Coefficients are reported in current UGX. Standard errors are adjusted for cluster robustness. We define an aggressive trim as that which results in a higher estimate; this may involve trimming observations from the bottom of the treatment group or from the top of the counterfactual.

Table 10
Between-village counterfactuals.

		Total Consumption	Total Net Assets	Total Productive Cash Inflows
Impact of Microenterprise Programs [b][ii]	Coefficient	15,164	13,517	9,057
	Error	19,131	9,470	
	p value	0.430	0.066 *	0.341
Impact of Transfer Programs [c][ii]	Coefficient	−31,299	14,696	−12,026
	Error	25,552	9,629	12,124
	p value	0.225	0.131	0.325

Notes: Outcomes are pooled across survey rounds. Coefficients are reported in current UGX. Only E1 is used as the counterfactual; this corresponds to the comparison sets ending in [ii].

Table 11
Internal rate of return under alternative assumptions.

	Variable name	
Investment per household (UGX at time of intervention)		571,147
Investment per household (constant 2016 USD)	I	215.36
Yearly consumption impact per capita (UGX at time of measurement)		26,061
Monthly consumption impact per household (constant 2016 USD)	c_m	3.84
Number of months between intervention and measurement date	m	21
Monthly rate of return (assuming persistent consumption effects)	τ (when r = 0%)	1.78%
Annualized rate of return		23.64%
Monthly rate of return (assuming diminishing consumption effects)	τ (when r = 3.17%)	0.23%
Annualized rate of return		2.78%

Notes: For the derivation of τ, see Appendix B: Rate of Return. All calculations are in constant USD; for the exchange rates used to define all monetary values in 2016 dollars, see Appendix A: Exchange Rates. The intervention date is defined as the weighted average transfer date; the date associated with the pooled measure is defined as the average date of the follow-up surveys; and the dates used to calculate rate r are defined as the dates of the two follow-up surveys. The value of investment equals the average cost of sub-arms A2 and B2 (see Table 2), weighted by the number of villages assigned to each sub-arm (see Fig. 1). The transformation of per capita to household values relies on the average household size reported in Table 3.

intervals expand and point estimates become more attenuated.

6.4. Sensitivity to assumptions about the persistence of returns

In Appendix B: Internal Rate of Return, we formalize social returns as a function of estimated consumption impacts and of the rate at which this impact diminishes. Using this formula, Table 11 shows that the rate of return to the microenterprise program is approximately 24% (in real and annualized terms, compounded monthly) if we assume that the estimated consumption effects from Table 5 persist. Meanwhile, if we assume that consumption effects fall in accordance with the point estimates from Table 13,[9] the rate of return is reduced to 3%. While it remains positive under this more pessimistic assumption, it may lie below the opportunity cost of capital (defined perhaps as the rate of return to other graduation programs as quantified by Banerjee et al. (2015); or as the rate of return to World Bank projects as reported by Herrera, 2005).

7. Discussion

7.1. Impacts

Consistent with Banerjee et al. (2015, 2017b), Bandiera et al. (2017), and Blattman et al. (2016), we found that an integrated microenterprise program targeting people in extreme poverty increased entrepreneurial activity and reduced poverty. Impacts on key economic outcomes appear significant, robust to multiple inference adjustments, and supported by consistent signals on subjective well-being and nutrition. The program seems cost-effective, with estimated gains offsetting the full costs of the program within roughly four years. This is based on outcomes that are pooled across follow-up rounds, which are statistically strongest because pooling tends to cancel out noise (McKenzie, 2012). If we disaggregate impacts into individual survey rounds, we see a possible attenuation in poverty effects over time, and rates of return fall substantially. We are therefore not able to speak confidently to the sustainability of gains. A further caveat is that negative spillovers could have somewhat biased our findings upward.

Like Karlan et al. (2017), we found that savings groups did little to encourage saving, but that they increased microenterprise activity and altered conditions in the community, especially with regards to the standing of women. Consistent with Banerjee et al. (2018), Berge et al. (2015), and Chowdhury et al. (2017), the program seems to lose impact when reduced to mere transfers. Combining transfers with a light-touch psychological intervention did tend to manipulate psychological outlook and to increase investment in productive (livestock) assets, but we do not see any indication of improvements consumption or cash inflows.

Consistent with Banerjee et al. (2017b) we do not find that cash transfers reduced labor market participation – on the contrary. That said, the estimated economic impacts of the pure cash transfers were very muted. At the time of measurement, we encounter some residual asset stock but no indications of increased cash income or consumption. Taken at face value, the results indicate that recipients dissaved and consumed a large share of their newly obtained assets very soon after the intervention, and that they did not invest the rest productively. This seems to vindicate the implementer's belief that without dedicated training and coaching, transfer recipients would struggle to maintain their newly received assets and derive sustained economic value from them. We cannot pinpoint what facet of the microenterprise program was critical and what importance should be attributed to coaching, business group formation, or other training modules. That said, our findings do not indicate that savings group formation alone explains the incremental impact of microenterprise programming. Like the data used by Haushofer and Shapiro (2018), ours indicate negative spillovers in cash villages.

The impacts of pure cash transfers must be interpreted with ample caution. Participants who were assigned to cash transfers chose to respond to the follow-up surveys at higher rates than participants in the pure control and microenterprise groups, which could have introduced

[9] The impacts associated with specific follow-up surveys are deflated using the corresponding rates in Appendix A: Exchange Rates.

attrition bias. Also, the pattern of low reported impacts on economic outcomes alongside high reported impacts on labor market participation raises the question if "motivated misreporting" may have played a role. Prior exposure to the cash transfer program could have conceivably led respondents in cash transfer villages to incorrectly assume that our survey enumerators might be involved in the selection of future transfer recipients, leading them to systematically under-report socioeconomic conditions with the objective of gaining future benefits. There is mixed support for this effect (Baird and Özler, 2012; Beegle et al., 2012; Martinelli and Parker, 2009; Moore et al., 2000). Motivated misreporting could have also been motivated by perceptions of fairness and equity. For example, there is a notable difference in the spillover effects that materialize in the data by Haushofer and Shapiro (2018), which tend negative; and those identified by Egger et al., (2019) in a very similar context, which are positive. One notable difference is that the first of these trials was designed like ours in that it targeted a random sub-sample of eligible households in cash villages with cash transfers; meanwhile, the latter targeted all eligible households, and we speculate that this could have been perceived as fairer by study participants. Beyond the patterns of results, we have no basis - not even anecdotes - that suggest differential misreporting. But we consider the possibility a real one; we urge enhanced attention to this issue in future studies and underline the merits of corroborating survey measures with ones that are not self-reported.

Overall, the results support the notion that integrated microenterprise programs in the spirit of ultra-poor graduation are sensible poverty reduction tools. While some caution is warranted, they also tend to support the more specific belief of the implementer that an integrated package, designed with market failures and internalities in mind, cannot be easily stripped of its components without adverse consequences.

7.2. Methodology

We explored an approach to safeguard against the data mining problem while maintaining the freedom to customize specifications to the data. In principle, this can avoid the epistemic cost of pre-analysis plans. It is possible to accomplish the same objective in other ways: for instance, data can be operationalized and models can be evaluated while treatment assignment remains blinded (Olken, 2015). That said, engaging in specification without access to treatment assignments can miss some important insights: for example, we had no intention to test for spillovers that are specific to cash transfer villages, but the specification curves forced us to contend with these. In the end, maintaining the option to customize specification choices to individual tests yielded few benefits because one econometric model ended up dominating others in our data set. This vindicates McKenzie's (2012) prescription to use ANCOVA in the analysis of economics field experiments. Following such contemporary norms, a skilled researcher may well have found the same results using pre-analysis plan.

Transparency and replicability

De-identified data, cleaning code, analysis code, scripts, surveys, and other supplementary materials are archived on the Open Science Framework (osf.io/mzrkx). The study was registered on the Registry for International Impact Evaluations (ID 52bb3799ccf6a).

Author contribution

RS and MS co-designed the study. AS took responsibility for the design of the behavioral intervention. MS supervised field work. RS designed and conducted the analysis. MS independently replicated key aspects of the analysis. RS wrote the paper.

Acknowledgments

This study could not have materialized without the extensive intellectual and managerial contributions of Rachel Proefke. We thank Sam Gant, who effectively led the local piloting and roll-out of the behavioral intervention. Other critical contributions were made by Dianne Calvi, Caroline Bernadi, Konstantin Zvereff, Winnie Auma, AJ Doty, Celeste Brubaker, and the team at Village Enterprise; by Dustin Davis, Agne Pupienyte, Harrison Pollock, Paola Elice, Samuel Rosenow, and the team at Innovations for Poverty Action; and by the team at the Independent Research and Evaluation Cell at BRAC Uganda. We benefitted from relevant conversations with Richard Tugume, Kate Orkin, Rob Garlick, Natalie Quinn, Berk Özler, Paul Sparks, Juan Camilo Villalobos, Kenneth Burnham, Uri Simonsohn, Ted Miguel, Don Green, Stefan Dercon, Michael Kremer, Bruce Wydick, and Martin Williams. We appreciate feedback from Dean Karlan, Jeremy Shapiro, Michael Cooke, and conference participants at the Buffett Institute, the World Bank, and the Centre for the Study of African Economies. Any errors are ours.

Appendix A. Exchange rates

All calculations are conducted in current Ugandan shillings (UGX). Where current USD numbers, 2016 USD numbers, and 2016 PPP USD numbers are reported, they are derived directly from UGX numbers, using UGX/USD midpoint rates from daily xe.com data for nominal rates; annual World Bank data for PPP rates; and monthly data from the US Bureau of Labor Statistics for USD inflation. The effective dates and corresponding rates used in this paper are as follows:

(a) The outset of the project is defined as the initial trial registration date, Dec 8, 2013. Applicable rates: 2520 (current USD terms), 2611 (2016 USD terms), 1049 (2016 USD PPP terms).
(b) The baseline date is defined as half way through the planned survey time frame (March 15, 2014). Applicable rates: 2520 (current USD terms), 2575 (2016 USD terms), 1063 (2016 USD PPP terms).
(c) The intervention date is defined as the UGX-weighted average transfer date (Aug 1, 2014). Applicable rates: 2613 (current USD terms), 2652 (2016 USD terms), 1056 (2016 USD PPP terms).
(d) The midline date is defined as half way through the planned midline survey time frame (Nov 15, 2015). Applicable rates: 3468 (current USD terms), 3528 (2016 USD terms), 1008 (2016 USD PPP terms).
(e) The pooled follow-up date is defined as half way through the planned survey time frame of both mid- and endline (May 15, 2016). Applicable rates: 3323 (current USD terms), 3340 (2016 USD terms), 1094 (2016 USD PPP terms).
(f) The endline date is defined as half way through the planned endline survey time frame (Nov 15, 2016). Applicable rates: 3556 (current USD terms), 3557 (2016 USD terms), 1146 (2016 USD PPP).

Appendix B. Internal rate of return

The internal rate of return τ aligns the value of investment I (that is made in period 0) with the present value of consumption impacts c (that accrue in all subsequent periods i):

$$I = \sum_{i=1}^{\infty} c_i \left(\frac{1}{1+\tau}\right)^i \tag{IV}$$

In measurement period m, consumption impact equals c_m. To quantify consumption impacts before and after period m, we assume that they diminish at constant rate r.

$$I = \sum_{i=1}^{\infty} c_m (1-r)^{i-m} \left(\frac{1}{1+\tau}\right)^i \tag{V}$$

Re-arranging:

$$\frac{I}{c_m}(1-r)^m = \sum_{i=1}^{\infty} \left(\frac{1-r}{1+\tau}\right)^i \tag{VI}$$

The right hand side is an infinite-horizon problem (Gordon and Shapiro, 1956) that can be simplified as follows[10]:

$$\frac{I}{c_m}(1-r)^m = \frac{1-r}{\tau+r} \tag{VII}$$

Solving for τ, we find:

$$\tau = \frac{c_m}{I}(1-r)^{1-m} - r \tag{VIII}$$

Appendix C. Participant flow

Table 12
Participant Flow over the Project's Life, by Sub-Arm and Cohort.

Sub-Arm	(1) Available Participant Slots				(2) Successful Baseline			
	Cohort #1	Cohort #2	Cohort #3	All	Cohort #1	Cohort #2	Cohort #3	All
A1	360	360	360	1,080	347	331	336	1,014
A2	420	420	420	1,260	404	384	391	1,179
B1	240	240	240	720	229	235	221	685
B2	280	280	280	840	266	265	260	791
C1	60	60	60	180	54	57	56	167
D1	168	168	168	504	156	155	152	463
D2	84	84	84	252	81	80	82	243
D3	84	84	84	252	78	81	78	237
E1	360	360	360	1,080	341	322	332	995
Total	2,056	2,056	2,056	6,168	1,956	1,910	1,908	5,774

Sub-arm	(3) Successful Midline					(4) Successful Endline				
	Cohort #1	Cohort #2	Cohort #3	All	Attrition(1)	Cohort #1	Cohort #2	Cohort #3	All	Attrition(1)
A1	316	302	321	939	7.40%	308	285	320	913	9.96%
A2	358	350	365	1,073	8.99%	354	335	370	1,059	10.18%
B1	215	219	211	645	5.84%	209	214	207	630	8.03%
B2	255	246	245	746	5.69%	249	230	245	724	8.47%
C1	43	54	53	150	10.18%	47	52	52	151	9.58%
D1	144	139	147	430	7.13%	138	136	145	419	9.50%
D2	78	78	78	234	3.70%	77	74	79	230	5.35%
D3	77	77	75	229	3.38%	75	72	76	223	5.91%
E1	314	304	315	933	6.23%	310	297	308	915	8.04%
Total	1,800	1,769	1,810	5,379	6.84%	1,767	1,695	1,802	5,264	8.83%

Notes: Available participant slots correspond to the numbers in Fig. 1. Only respondents who consented to and completed to baseline survey were recruited into the study. Of the resulting study population, follow-ups were successful with 93% and 91% of respondents in the two respective follow-up surveys. (1) Attrition is defined as the share of baseline survey participants for whom the corresponding follow-up survey was unsuccessful.

10 $\sum_{i=1}^{\infty}\mu^i = \mu+\mu^2+\ldots+\mu^{\infty}$ and $\mu\sum_{i=1}^{\infty}\mu^i = \mu^2+\ \ldots+\mu^{\infty}$. Therefore, if $-1<\mu<1$ then it follows that $\sum_{i=1}^{\infty}\mu^i - \mu\sum_{i=1}^{\infty}\mu^i = \mu$. Re-arranging: $\sum_{i=1}^{\infty}\mu^k = \frac{\mu}{1-\mu}$. Under the assumption that $-1<\mu=\frac{1-r}{1+\tau}<1$, it follows that $\sum_{i=1}^{\infty}\left(\frac{1-r}{1+\tau}\right)^i = \frac{\frac{1-r}{1+\tau}}{1-\frac{1-r}{1+\tau}} = \frac{1-r}{\tau+r}$.

Appendix D. Additional figures and tables

This Appendix shows figures and tables that were mentioned but not included in the paper. For the complete set, consult the Online Appendix of Specification Curves and the Online Appendix of Tables.

Table 13
Impact of Microenterprise Programs on Poverty Indicators.

Follow-up Round	First	Second	Pooled	First	Second	Pooled	First	Second	Pooled
	Total Consumption			Total Assets			Total Productive Cash Inflows		
Coefficient	27,526	18,859	26,061	20,189	10,570	16,343	20,447	8,889	13,483
Error	14,617	12,434	11,248	5,374	5,552	5,449	9,738	8,185	6,747
Effect/mean	0.04	0.03	0.04	0.17	0.09	0.14	0.15	0.07	0.10
Effect/sd	0.05	0.05	0.07	0.14	0.07	0.12	0.08	0.04	0.07
p value	0.062 *	0.132	0.022 **	0.000 ***	0.059 *	0.003 ***	0.038 **	0.279	0.048 **
q value	0.094 *	0.142	0.055 *	0.003 ***	0.094 *	0.021 **	0.076 *	0.218	0.087 *
N	4,750	4,655	4,906	4,750	3,598	3,796	3,901	3,815	4,021
	Food and Beverage Consumption			Livestock Assets			Net Cash Inflows from Farming		
Coefficient	28,334	15,898	25,180	13,134	8,182	10,584	−2,812	4,359	1,409
Error	12,875	10,088	9,381	3,092	2,954	2,657	2,934	2,596	2,161
Effect/mean	0.05	0.03	0.05	0.22	0.16	0.19	−0.26	0.23	0.10
Effect/sd	0.07	0.05	0.08	0.14	0.10	0.14	−0.03	0.06	0.02
p value	0.029 **	0.117	0.008 ***	0.000 ***	0.006 ***	0.000 ***	0.340	0.095 *	0.515
q value	0.069 *	0.133	0.029 **	0.002 ***	0.029 **	0.002 ***	0.252	0.118	0.304
N	4,750	4,655	4,906	4,750	3,718	4,906	3,901	4,801	4,021
	Recurring Consumption			Durable Assets			Income from Other Self-Employment		
Coefficient	−1,690	−1,411	−1,402	6,531	1,996	4,440	17,784	6,700	11,862
Error	2,056	2,377	1,917	2,510	2,936	2,452	5,477	5,381	4,361
Effect/mean	−0.02	−0.02	−0.02	0.12	0.03	0.08	0.28	0.12	0.20
Effect/sd	−0.02	−0.02	−0.02	0.09	0.02	0.06	0.13	0.05	0.11
p value	0.413	0.554	0.466	0.010 **	0.498	0.072 *	0.001 ***	0.215	0.007 ***
q value	0.275	0.308	0.296	0.033 **	0.303	0.099 *	0.013 **	0.190	0.029 **
N	4,916	4,811	5,073	3,901	3,695	3,901	3,796	4,655	3,916
	Infrequent Consumption			Net Financial Position			Income from Paid Employment		
Coefficient	1,393	4,638	2,839	506	1,905	1,238	−630	−2,217	−1,088
Error	3,260	2,443	2,605	701	707	572	2,416	3,622	2,681
Effect/mean	0.02	0.07	0.04	0.57	1.46	1.11	−0.02	−0.06	−0.03
Effect/sd	0.01	0.06	0.04	0.03	0.10	0.09	−0.01	−0.02	−0.01
p value	0.670	0.060 *	0.278	0.472	0.008 ***	0.032 **	0.795	0.542	0.686
q value	0.378	0.094 *	0.218	0.296	0.029 **	0.070 *	0.448	0.308	0.378
N	3,796	3,718	3,916	3,901	3,815	4,021	4,750	3,718	3,916

Notes: Coefficients are reported in current Ugandan Shillings per capita. Flow variables (consumption and income) are yearly. Effect/mean is the coefficient divided by the average value of the counterfactual. Effect/sd is the coefficient divided by the standard error of the counterfactual. Estimates pertain to coefficient β in the preferred specification for so-called clustered comparisons. Intent-to-treat assignment T is coded to the value one among households in set $A2 \cup B2$ and to the value zero in set $A1 \cup B1 \cup C1 \cup D1 \cup E1$. Errors are adjusted for cluster robustness. Food and beverage consumption is measured using regional market prices, i.e., median prices by region (East or West) and survey wave (Baseline, First Follow-up, Second follow-up). Recurring and infrequent consumption items are valued using self-reported prices. Total consumption is the sum of the three consumption sub-composites. Livestock assets are valued using local median prices. Durable assets are valued using self-reported prices and include the capitalized expenditures on home improvements; no depreciation is applied. The net financial position corresponds to the outstanding value of loans minus the value of savings. Total assets are the sum of the three asset sub-composites. Net cash inflows from farming are the cash sales of perennial crops by the household, plus the cash sales of crops grown by the household and with business partners over the last short and long season preceding the surveys, plus the cash sales of animals and animal products (with all sales being valued at regional market prices), minus all cash expenditures on farming and livestock. Income from other self-employment and paid employment are derived directly from the survey in cash terms. Total productive cash inflows are the aggregate of thee cash inflow sub-composites. For more information, consult the survey forms, data, and code.

Table 14
Impact of Transfer Programs on Schooling.

Follow-up Round		First Follow-up		Second Follow-up	
		Specification 1	Specification 2	Specification 1	Specification 2
Enrolled in and Attending School	Odds ratio	0.950	0.871	1.324	1.182
	Error	0.135	0.147	0.162	0.205
	p value	0.716	0.414	0.022 **	0.336
	q value	1.000	1.000	0.537	1.000
	N	7,760	5,097	9,818	5,212
Repeated Year	Odds ratio	0.959	0.977	0.878	0.882
	Error	0.104	0.105	0.090	0.107
	p value	0.697	0.830	0.202	0.300
	q value	1.000	1.000	1.000	1.000
	N	6,497	4,081	7,710	3,971
Days worked last Month	Coefficient	−1.334	−1.139	0.498	−0.688
	Error	3.559	3.386	2.147	2.401
	p value	0.708	0.737	0.817	0.775
	q value	1.000	1.000	1.000	1.000
	N	7,889	6,192	9,819	5,291
School Days Missed last Month	Coefficient	−0.095	−0.179	−0.227	−0.062
	Error	0.124	0.137	0.156	0.178
	p value	0.443	0.193	0.147	0.729
	q value	1.000	1.000	1.000	1.000
	N	6,502	3,973	7,573	3,868

Notes: Specification 1 uses simple regression. Specification 2 uses the preferred model for so-called clustered comparisons. In the case of binary outcomes, logistic regression is applied, estimates are presented as odds ratios, and no pooled follow-up round is created. Errors are adjusted for cluster robustness. Intent-to-treat assignment T is coded to the value one among households in set D2∪D3 and to the value zero in set A1∪B1∪C1∪D1∪E1. The unit of observation is not the household, but the individual household member; only members aged 5 through 17 are included.

Table 15
Impact of Transfer Programs on Financial Position.

Follow-up Round		First Follow-up		Second Follow-up		Pooled Follow-ups	
		Specification 1	Specification 2	Specification 1	Specification 2	Specification 1	Specification 2
Savings	Coefficient	1,431	1,499	2,227	2,392	1,811	1,887
	Error	1,190	1,090	1,504	1,451	1,208	1,085
	p value	0.231	0.171	0.141	0.101	0.136	0.084 *
	q value	0.301	0.265	0.229	0.198	0.229	0.180
	N	3,560	2,840	3,481	2,764	3,661	2,916
Loans	Coefficient	−1,170	−2,013	−821	−1,369	−939	−1,648
	Error	529	543	618	670	485	491
	p value	0.029 **	0.000 ***	0.186	0.043 **	0.055 *	0.001 ***
	q value	0.114	0.006 ***	0.268	0.137	0.148	0.009 ***
	N	3,560	2,840	3,481	2,764	3,661	2,916

Notes: Coefficients are reported in current UGX per capita. Specification 1 uses simple linear regression. Specification 2 uses the preferred model for so-called clustered comparisons. Intent-to-treat assignment T is coded to the value one among households in set D2∪D3 and to the value zero in set A1∪B1∪C1∪D1∪E1. Errors are adjusted for cluster robustness. Intent-to-treat assignment T is coded to the value one among households in set A1∪B1∪C1∪D1 and to the value zero in set E1. Errors are adjusted for cluster robustness. Loans are defined as the estimated outstanding value.

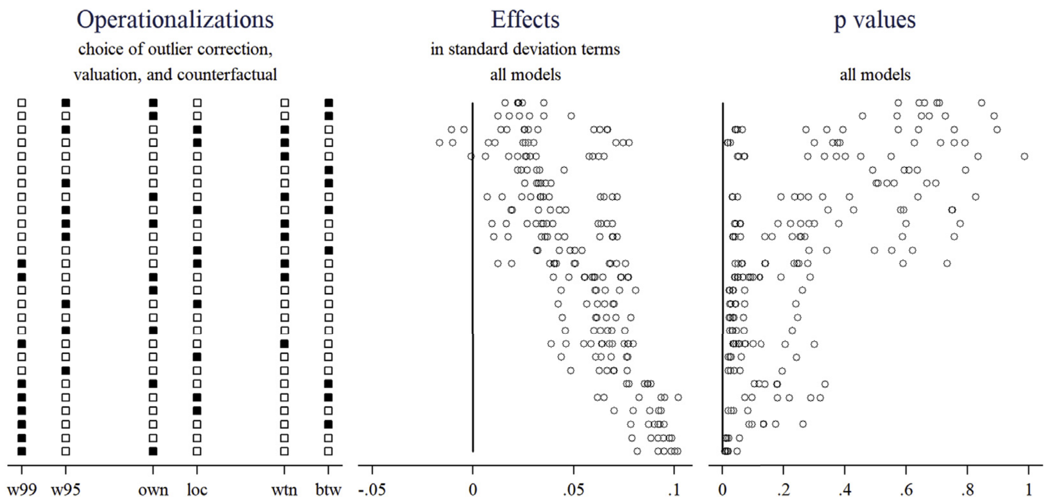

Fig. 8. Impact of Microenterprise Programs on Consumption (Specification Curve).

Table 16
Impact of Transfer Programs on Labor

Follow-up round		First Follow-up		Second Follow-up	
		Specification 1	Specification 2	Specification 1	Specification 2
Active in Labor Force	Odds ratio	1.150	1.222	1.260	1.361
	Error	0.153	0.169	0.120	0.152
	p value	0.292	0.146	0.015 **	0.006 ***
	q value	0.502	0.264	0.051 *	0.028 **
	N	9,609	7,418	10,482	7,449
Active in Microenterprise	Odds ratio	1.278	1.317	1.402	1.550
	Error	0.151	0.157	0.122	0.194
	p value	0.038 **	0.021 **	0.000 ***	0.000 ***
	q value	0.075 *	0.057 *	0.002 ***	0.004 ***
	N	9,611	6,061	10,500	6,046
Active as Employee or Day Laborer	Odds ratio	0.945	1.011	0.999	1.033
	Error	0.124	0.127	0.140	0.127
	p value	0.666	0.933	0.994	0.793
	q value	0.999	1.000	1.000	1.000
	N	9,619	7,434	10,514	6,053
Active in more than one Livelihood	Odds ratio	0.981	1.058	0.883	0.915
	Error	0.108	0.121	0.114	0.126
	p value	0.860	0.622	0.337	0.517
	q value	1.000	0.999	0.508	0.871
	N	9,621	7,436	10,517	7,478

Notes: Specification 1 uses simple regression. Specification 2 uses the preferred model for so-called clustered comparisons. In the case of binary outcomes, logistic regression is applied, estimates are presented as odds ratios, and no pooled follow-up round is created. Errors are adjusted for cluster robustness. Intent-to-treat assignment T is coded to the value one among households in set D2∪D3 and to the value zero in set A1∪B1∪C1∪D1∪E1. The unit of observation is not the household, but the individual household member; only members aged 15 and over are included. Microenterprise activity implies either farm-based or non-farm-based self-employment. Labor force participation implies microenterprise activity or paid work activity.

Bibliography

Abadie, A., Athey, S., Imbens, G., Woolridge, G., 2017. When Should You Adjust Standard Errors for Clustering? (NBER Working Paper No. 24003).

Anderson, M.L., 2008. Multiple inference and gender differences in the effects of early intervention: a reevaluation of the Abecedarian, Perry preschool, and early training projects. J. Am. Stat. Assoc. 103 (484), 1481–1495.

Baird, S., Özler, B., 2012. Examining the reliability of self-reported data on school participation. J. Dev. Econ. 98 (1), 89–93.

Bandiera, O., Burgess, R., Das, N., Gulesci, S., Rasul, I., Sulaiman, M., 2017. Labor markets and poverty in village economies. Q. J. Econ. 811–870.

Banerjee, A., Duflo, E., Chattopadhyay, R., Shapiro, J., 2009. Targeting Efficiency: How Well Can We Identify the Poorest of the Poor? mimeo.

Banerjee, A., Duflo, E., Chattopadhyay, R., Shapiro, J., 2017a. The Long Term Impacts of a "Graduation" Program: Evidence from West Bengal. Massachusetts Institute of Technology, pp. 1–25 (September).

Banerjee, A., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Parienté, W., Udry, C., 2015. A multifaceted program causes lasting progress for the very poor: evidence from six countries. Science 348 (6236), 1260799.

Banerjee, A., Hanna, R., Kreindler, G., Olken, B.A., 2017b. Debunking the Stereotype of the Lazy Welfare Recipient: Evidence from Cash Transfer Programs Worldwide (Faculty Research Paper Series No. RWP15-076).

Banerjee, A., Karlan, D., Osei, R.D., Trachtman, H., Udry, C., 2018. Unpacking a Multi-Faceted Program to Build Sustainable Income for the Very Poor (No. WP-18-04).

Bastagli, F., Hagen-Zanker, J., Harman, L., Barca, V., Sturge, G., Schmidt, T., Pellerano, L., 2016. Cash Transfers: what Does the Evidence Say? A Rigorous Review of Programme Impact and of the Role of Design and Implementation Features. Overseas Development Institute, p. 300 (July).

Bauer, P., 1972. Dissent on Development. Harvard University Press, Cambridge, MA.
Beegle, K., De Weerdt, J., Friedman, J., Gibson, J., 2012. Methods of household consumption measurement through surveys: experimental results from Tanzania. J. Dev. Econ. 98 (1), 3–18.
Benhassine, N., Devoto, F., Duflo, E., Dupas, P., Pouliquen, V., 2015. Turning a shove into a nudge? A "labeled cash transfer" for education. Am. Econ. J. Econ. Policy 7 (3), 1–48.
Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Royal Stat. Soc. B 57 (1), 289–300.
Berge, L.I.O., Bjorvatn, K., Tungodden, B., 2015. Human and financial capital for microenterprise development: short-term and long-term evidence from a field experiment in Tanzania. Manag. Sci. 61 (4), 707–722.
Blattman, C., Fiala, N., Martinez, S., 2018. The Long Term Impacts of Grants on Poverty: 9-year Evidence from Uganda's Youth Opportunities Program (NBER Workine Paper Series No. 24999).
Blattman, C., Green, E.P., Jamison, J., Christian, M., Annan, J., Carlson, N., et al., 2016. The returns to microenterprise support among the ultrapoor: a field experiment in postwar Uganda. Am. Econ. J. Appl. Econ. 8 (2), 35–64.
Bold, T., Kimenyi, M., Mwabu, G., Ng'ang'a, A., Sandefur, J., 2013. Scaling-up what Works: Experimental Evidence on External Validity in Kenyan Education (Working Paper Series No. 321). Center for Global Development.
Burnham, K.P., Anderson, D.R., 2004. Multimodel inference: understanding AIC and BIC in model selection. Sociol. Methods Res. 33 (2), 261–304.
Burnham, K.P., Anderson, D.R., 2010. Model Selection and Multi-Model Inference: A Practical Information-Theoretical Approach. Springer, New York.
Cartwright, N., Hardie, J., 2012. Evidence-Based Policy: A Practical Guide to Doing it Better. Oxford University Press, USA.
Casey, K., Glennerster, R., Miguel, E., 2012. Reshaping institutions: evidence on aid impacts using a preanalysis plan. Q. J. Econ. 127 (4), 1755–1812.
Chowdhury, R., Collins, E., Ligon, E., Sulaiman, M., 2017. Valuing Assets provided to Low-Income Households in South Sudan. mimeo.
Cialdini, R., Trost, M., 1998. Social influence: Social norms, conformity and compliance. In: The Handbook of Social Psychology, 2.
Clyde, M., 2003. Model averaging. In: Press, S.J., Chib, S. (Eds.), Subjective and Objective Bayesian Statistics. John Wiley & Sons, Inc, Hoboken, NJ, USA.
Cohen, G.L., Sherman, D.K., 2014. The psychology of change: self-affirmation and social psychological intervention. Annu. Rev. Psychol. 65 (1), 333–371.
De Mel, S., McKenzie, D., Woodruff, C., 2012. One-time transfers of cash or capital have long-lasting effects on microenterprises in Sri Lanka. Science 335 (6071), 962–966.
Deaton, A., 1997. The Analysis of Household Surveys: A Microeconometric Approach to Development Policy. World Bank Publishing Group, Washington DC.
Deaton, A., 2010. Instruments, randomization, and learning about development. J. Econ. Lit. 48 (2), 424–455.
DellaVigna, S., 2009. Psychology and economics: evidence from the field. J. Econ. Lit. 47 (2), 315–372.
Duflo, E., 2012. Human values and the design of the fight against poverty. In: Conference Notes: Tanner Lectures.
Easterly, W., 2007. The White Man's Burden: Why the West's Efforts to Aid the Rest Have Done So Much Ill and So Little Good. Penguin Group.
Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., 2004. Least angle regression. Ann. Stat. 32 (2), 407–499.
Egger, E., Haushofer, J., Miguel, E., Niehaus, P., Walker, M., 2019. General Equilibrium Effects of Cash Transfers: Experimental Evidence from Kenya (NBER Working Paper No. 16205).
Fafchamps, M., McKenzie, D., Quinn, S., Woodruff, C., 2014. Female microenterprises and the fly-paper effect: evidence from a randomized experiment in Ghana. J. Dev. Econ. 106, 211–226.
Festinger, L., 1957. A Theory of Cognitive Dissonance. Stanford University Press, Stanford.
Fiala, N., 2018. Returns to microcredit, cash grants and training for male and female microentrepreneurs in Uganda. World Dev. 105, 189–200.
Gash, M., Odell, K., 2013. The Evidence-Based Story of Savings Groups: A Synthesis of Seven Randomized Control Trials (SEEP Working Paper).
Gollwitzer, P.M., 1999. Implementation intentions. Am. Psychol. 54 (7), 493–503.
Gordon, M.J., Shapiro, E., 1956. Capital equipment analysis: the required rate of profit. Manag. Sci. 3 (1), 102–110.
Hall, C.C., Zhao, J., Shafir, E., 2014. Self-affirmation among the poor: cognitive and behavioral implications. Psychol. Sci. 25 (2), 619–625.
Haushofer, J., Shapiro, J., 2018. The Long-Term Impacts of Unconditional Cash Transfers: Experimental Evidence from Kenya. mimeo.
Herrera, S., 2005. The Economic Rate of Return on World Bank Projects. Washington, DC.
Hoeting, J., 2002. Methodology for Bayesian model averaging: an update. In: International Biometric Conference, pp. 231–240.
Hoeting, J., Madigan, D., Raftery, A.E.A., Volinsky, C.T., 1999. Bayesian model averaging: a tutorial. Stat. Sci. 14 (4), 382–401.
Ioannidis, J.P.A., 2005. Why most published research findings are false. PLoS Med. 2 (8), 0696–0701.
Jensen, R., 2010. The (perceived) returns to education and the demand for schooling. Q. J. Econ. 125 (2), 515–548.
Karlan, D., Knight, R., Udry, C., 2015. Consulting and capital experiments with microenterprise tailors in Ghana. J. Econ. Behav. Organ. 118, 281–302.
Karlan, D., McConnell, M., Mullainathan, S., Zinman, J., 2010. Getting to the Top of Mind: How Reminders Increase Saving (NBER Working Paper No. 16205).
Karlan, D., Savonitto, B., Thuysbaert, B., Udry, C., 2017. Impact of savings groups on the lives of the poor. Proc. Natl. Acad. Sci. 114 (12), 3079–3084.
Kling, J.R., Liebman, J.B., Katz, L.F., 2007. Experimental analysis of neighborhood effects. Econometrica 75 (1), 83–119.
Lee, D.S., 2018. Training, wages, and sample selection: estimating sharp bounds on treatment effects. Rev. Econ. Stud. 76 (May), 1071–1102.
Ljungqvist, L., 1993. Economic underdevelopment: the case of a missing market for human capital. J. Dev. Econ. 40 (2), 219–239.
Lockwood, P., Kunda, Z., 1997. Superstars and me: Predicting the impact of role models on the self. J. Personal. Soc. Psychol. 73 (1), 91–103.
Martinelli, C., Parker, S.W., 2009. Deception and misreporting in a social program. J. Eur. Econ. Assoc. 7 (4), 886–908.
McKenzie, D., 2012. Beyond baseline and follow-up: the case for more T in experiments. J. Dev. Econ. 99 (2), 210–221.
McKenzie, D., 2017. Identifying and spurring high-growth entrepreneurship: experimental evidence from a business plan competition. Am. Econ. Rev. 107 (8), 2278–2307.
Meyer, B.D., Sullivan, J.X., 2003. Measuring the Well-Being of the Poor Using Income and Consumption (NBER Working Paper No. 9760).
Moore, J.C., Stinson, L.L., Welniak, E.J., 2000. Income measurement error in surveys: a review. J. Off. Stat. 16 (4), 331–361.
Moyo, D., 2009. Dead Aid. Farrar, Straus and Giroux, New York.
O'Donoghue, T., Rabin, M., 1999. Doing it now or later. Am. Econ. Rev. 89 (1), 103–124.
Oettingen, G., 2000. Expectancy effects on behavior depend on self-regulatory thought. Soc. Cognit. 18 (2), 101–129.
Olken, B.A., 2015. Promises and perils of pre-analysis plans. J. Econ. Perspect. 29 (3), 61–80.
Pritchett, L., Sandefur, J., 2014. Context matters for size: why external validity claims and development practice do not mix. J. Glob. Dev. 4 (2), 161–197.
Rogers, T., Milkman, K.L., 2016. Reminders through association. Psychol. Sci. 27 (7), 973–986.
Schreiner, M., 2011. Simple Poverty Scorecard Poverty-Assessment Tool Uganda. mimeo.
Shapiro, J., 2017. Benchmarking Development Programs: A Preference-Based Approach.
Simonsohn, U., Simmons, J.P., Nelson, L.D., 2015. Specification Curve: Descriptive and Inferential Statistics on All Reasonable Specifications (SSRN Working Paper No. 2694998).
Stiglitz, J.E., 1989. Markets, market failures, and development. Am. Econ. Rev. 79 (2), 196–203.
Thaler, H., 1999. Mental accounting matters. J. Behavior. Des. Making 12, 183–206.
Wagenmakers, E.-J., Farrell, S., 2004. AIC model selection using Akaike weights. Psychon. Bull. Rev. 11 (1), 192–196.
World Bank, 2019. The Atlas of Social Protection: Indicators of Resilience and Equity (ASPIRE Database). World Bank Group, Washington, DC. Retrieved from: https://web.archive.org/web/20190129141542/http://datatopics.worldbank.org/aspire/indicator_glance.
Wydick, B., 2018. When Are Cash Transfers Transformative? (CEGA Working Paper No. 069).